TRANSACTIONS

OF THE

AMERICAN INSTITUTE

OF

ELECTRICAL ENGINEERS

VOLUME 78
1959

PART I. COMMUNICATION AND ELECTRONICS

PUBLISHED BY THE
American Institute of Electrical Engineers
33 WEST 39th STREET
NEW YORK 18, N. Y., U. S. A.

Copyright 1960 by the American Institute of Electrical Engineers

Printed in U. S. A. by The Mack Printing Company, Easton Pa.

Preface

The AIEE *Transactions* for 1959 (volume 78) is published in three parts: Part I. Communication and Electronics; Part II. Applications and Industry; and Part III. Power Apparatus and Systems. (Part III is divided into two parts: III–A, consisting of pages 1–976; III–B, consisting of pages 977–1870. The Index for both parts appears at the end of III–B.) The papers in each of the three parts are classified according to subject matter as follows:

Part I. Communication and Electronics

Aircraft and Flight Test Instrumentation
Basic Sciences
Broadcasting
Communication Switching Systems
Communication Theory
Computing Devices
Data Communication
Electrical Insulation
Electrical Techniques in Medicine and Biology
Electronic and High Frequency Instruments
Electronics
Indicating and Integrating Instruments
Magnetic Amplifiers
Nucleonic and Radiation Instrumentation
Nucleonics
Radio Communication Systems
Recording and Controlling Instrumentation
Semiconductor Metallic Rectifiers
Solid State Devices
Special Instruments and Auxiliary Apparatus
Telegraph Systems
Telemetering
Wire Communication Systems

Part II. Applications and Industry

Aero-Space Transportation
Chemical Industry
Domestic and Commercial Applications
Electric Heating
Electric Welding
Feedback Control Systems
General Industry Applications
Industrial and Commercial Power Systems
Industrial Control
Industrial Power Rectifiers
Land Transportation
Marine Transportation
Metal Industry
Mining Industry
Petroleum Industry
Production and Application of Light

Part III. Power Apparatus and Systems

Insulated Conductors
Power Generation
Power System Communications
Protective Devices
Relays
Rotating Machinery
Substations
Switchgear
System Engineering
Transformers
Transmission and Distribution

Research
Safety
Standards

Each part has been indexed separately in the back of that particular part. The three parts are not cross-referenced; hence the user should determine first whether the subject matter of the paper desired is in the field of communication and electronics, applications and industry, or power apparatus and systems. Papers are listed in the subject index under several key words in the titles. The original numbers assigned to the papers are given in the author index. Volume 78 contains the technical papers and related discussions presented at these meetings:

1. Winter General Meeting, New York, N. Y., February 1–6, 1959.
2. Southeast and South Central District Meeting, Atlanta, Ga., April 8–10, 1959.
3. East Central District Meeting and Rubber and Plastics Conference, Akron, Ohio, April 22–24, 1959.
4. Empire District Meeting, Syracuse, N. Y., April 29–May 1, 1959.
5. Middle Eastern District Meeting, Baltimore, Md., May 19–21, 1959.
6. Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959.
7. AIEE-IRE Nonlinear Magnetics and Magnetic Amplifiers Conference, Washington, D. C., September 23–25, 1959.
8. Fall General Meeting, Chicago, Ill., October 11–16, 1959.

Statements and opinions given in the papers and discussions published in *Transactions* are the expressions of the contributors for which the American Institute of Electrical Engineers assumes no responsibility.

Vol.
Pt. 1
1959

Communication and Electronics Index for 1959

.78
i
59

1. Technical Subject Index

A-C Magnets from Simple Models, Inductance of. Douglas, Voith....562–8; disc. 568

A-C Resistance of Solid Magnetic Wires, The. Gohar....912–16

A-C Shocks of Varying Parameters Affecting the Heart. Kouwenhoven, Knickerbocker, Chesnut, Milnor, Sass....163–7; disc. 167

Aging Calculations, Mathematics of Insulation-. Whitman, Whitman....308–16; disc. 316

Aircraft Electric Systems, Transistor-Magnetic Control Circuits for. Pratt....643–50

Algebra of Sequential Logic—Matrix Logic III, Matrix. Schubert....1074–9

Algebraic Matrix Equations, Electronic Network Synthesis of Linear. Horn, Honnell....1028–32

Allentown-Bethlehem, Pa., City-Wide Personal Signaling at. Kraus....52–5

Alloy Junction Transistors Above Rated Conditions, Effects of Operation of Germanium. Spradlin....346–52

Alternating Magnetic Fields, Conducting Sphere in. Poritsky....937–42

Analog Conversion, Magnetic Amplifier Binary-to-. Danylchuk, Katz....909–12

Analog Data, Measurements of Narrow-Band Noise on Telephone Facilities in Connection with. Edson, Froehlich, Townley....83–6

Analog Division Circuit. McMurray....606–12

Analog of Heat Flow in Power Transistors, An Electric. Reese, Grannemann, Durant....640–3

Analog Simulations, Application of Finite Integral Transforms to. Karplus, Stephens....152–5

Analogs and Duality, Electromechanical Impedance. Swift....81–3

Analysis and Design of a Transistor Linear-Delay Circuit. Nanavati....577–80

Analysis of Magnetic Amplifiers Using a Resistive Reactor Model. Frankenthal....850–7; disc. 857

Analysis of Magnetic Amplifiers With Diodes. Johannessen....485–504

Analysis of Magnetic Amplifiers Without Diodes. Johannessen....471–85

Analysis of Series-Connected Saturable Reactor with Capacitive Loading and Finite Control Resistance by Use of Difference Equations. Bourne, Salihi....461–71

Analysis of the Effects of an Electric Field on the Charging of Fine Particles, A Theoretical. Murphy, Adler, Penney....318–26

Annular-Plate Capacitors, Subarea Determination of the Capacitance of Cocentric. Reitan, Higgins....1002–05

Anomaly in Residual Polarization of $BaTiO_3$ Ceramics, On the. Saito, Yamanaka....70–6

Antenna, The Design of the Corner-Reflector. Neff, Tillman....293–5

Apparatus as Compared with Insulation Material Tests, Functional Life Tests of. Manning....1107–10; disc. 1110

Application of a New Carrier Telegraph System. Collins, Pitts....1009–16; disc. 1016

Application of Finite Integral Transforms to Analog Simulations. Karplus, Stephens....152–5

Application of Telegraph Techniques in Data Transmission. Boggs, Boughtwood....336–40

Approach to the Study of Automatic Target Tracking, An. Pode....806–11

Approach to Training Telephone Engineers, A New. Burnett, Adams....572–6; disc. 576

Artificial Respiration Techniques, A Comparison of Mouth-to-Mouth and Manual. Gordon, Frye....86–91

Assurance Program, Quality. Hulnick, Harding, Rowinski....326–35

Asynchronous Multiplexing. Taylor....1054–62

Automatic Programing Techniques for Solving Engineering Problems, The Use of. Carleton, Chackan, Martin....596–601; disc. 601

Automatic Regulators with Self-Balancing Magnetic Amplifiers. Geyger....433–8

Automatic Target Tracking, An Approach to the Study of. Pode....806–11

Automatic Teletypewriter Switching System, A Small. Shimmin, Vanderlippe, Whitman....959–64

B

Basic Concepts of Multidimensional Space Filters. Kron....554–61

(Basic Sciences) A New Direct Matrix Inversion Method. Shipley, Coleman....568–70; disc. 570

(Basic Sciences) Conducting Sphere in Alternating Magnetic Fields. Poritsky....937–42

(Basic Sciences) Criteria and Tests for Realizability of the Inductance Matrix. Tokad, Reed....924–6

(Basic Sciences) Current Distribution in the Cylindrical Source Plane-Electrode Configuration. Comstock, Williams....252–6

(Basic Sciences) Eddy-Current Losses in Solid and Laminated Iron. Agarwal....169–79; disc. 179

(Basic Sciences) Effects of Ultrahigh Temperature on Magnetic Properties of Core Materials. Pasnak, Lundsten....1033–8; disc. 1038

(Basic Sciences) Electromechanical Impedance: Analogs and Duality. Swift....81–3

(Basic Sciences) Electronic Network Synthesis of Linear Algebraic Matrix Equations. Horn, Honnell....1028–32

(Basic Sciences) Inductance of A-C Magnets from Simple Models. Douglas, Voith....562–8; disc. 568

(Basic Sciences) Mixed Transcendental and Polynomial Transforms. Smith....786–95

(Basic Sciences) Recurrence Relations in the Solution of a Certain Class of Nonlinear Systems. Wolf....830–4; disc. 834

(Basic Sciences) Some Multihorned Dilemmas in the Magnetic Field. Avcin....1087–95

(Basic Sciences) Spectral Output of Piecewise Linear Nonlinearity. Smith....543–9

(Basic Sciences) Statistical Spectral Output of Power Law Nonlinearity. Smith....535–43

(Basic Sciences) Subarea Determination of the Capacitance of Cocentric Annular-Plate Capacitors. Reitan, Higgins....1002–05

(Basic Sciences) The A-C Resistance of Solid Magnetic Wires. Gohar....912–16

(Basic Sciences) The Characteristic Impedance and Phase Velocity of a Shielded Helical Transmission Line. Kirschbaum....444–50

(Basic Sciences) The Computation of Correlation and Spectral Functions by Orthogonal Filtering. Gilbert....954–9

(Basic Sciences) The Effect of Flux Distribution on Iron Losses. Schindler....1069–74

(Basic Sciences) The General Form of Some Common Network Theorems. Adamson, El-Sobki....991–5; disc. 995

(Basic Sciences) The Optimum Transmission-Line Pulse Transformer. Young, Schatz, Woodford....220–3

(Basic Sciences) The Transient Response of Tapered Transmission Lines. Young, Schatz, Woodford....223–8; disc. 228

(Basic Sciences) Two-Oscillator Scan Stability. Joseph....1022–8

$BaTiO_3$ Ceramics, On the Anomaly in Residual Polarization of. Saito, Yamanaka....70–6

Bethlehem, Pa., City-Wide Personal Signaling at Allentown-. Kraus....52–5

Binary-to-Analog Conversion, Magnetic Amplifier. Danylchuk, Katz....909–12

Bridge for Surface Temperature Measurement, An Impedance. Mouly....388–92; disc. 392

Bridge Magnetic Amplifier, Theory of 3-Phase. Storm, Flairty:
  Part I....878–95

Brightness Intensification in Fluoroscopic Image Intensifiers, Standards for Measurement of. Lusby....504–05; disc. 506

Burial Wire, The Design and Manufacture of Direct. Robb, Roberts....662–6; disc. 666

C

Cable, Design of Polyethylene-Insulated Multipair Telephone. Windeler....736–9

Cable, More About Nonarmored Submarine. Lawton....367–9

Cable Sheath for Telephone Exchange Cable, Manufacture of Composite. Hughes....650–4

Cable Simulation Networks, Synthesis of. DeMonte....682–6

Cables at Voice and Carrier Frequencies, Transmission Properties of Polyethylene-Insulated Telephone. Eager, Jachimowicz, Kolodny, Robinson....618–39; disc. 639

Calculations, Mathematics of Insulation-Aging Whitman, Whitman....308–16; disc. 316

Calibration of Precision Voltage Dividers, A Method for. Pinckney....182–5

Calling System, The 83B1 Teletypewriter Selective-. Smith, Votaw, Whitman....119–24

Capacitance of Cocentric Annular-Plate Capacitors, Subarea Determination of the. Reitan, Higgins....1002–05

Capacitances in Magnetic Amplifiers, The Winding. Johansen....702–07

Capacitive Loading and Finite Control Resistance by Use of Difference Equations, Analysis of Series-Connected Saturable Reactor with. Bourne, Salihi....461–71

Capacitively Coupled Magnetic Amplifiers. Collins....707–12

Carbon-Film Resistors after 12,000 Hours of Operation, Life Characteristics of. Braner, Easterday..201–07

Carrier System, The K24A Syncroplex Telephone. Coetsee, Curtis, Halina....1044–54; disc. 1054

Carrier Systems Up to 156 Kc, A Transposition System for. Kirkland....612–14

Carrier Telegraph System, Application of a New. Collins, Pitts....1009–16; disc. 1016

Carrier Telegraph System, Design Features of a New Frequency Shift. Stephens....927–34

Carrier Telegraph Terminal, A Transistorized 20-Channel. Grybowski, Vieth....260–5

CCITT Recommendations for Multichannel Radio Relays and White Noise. Parry....107–17

Central Office Area, Indialing to PBX Extensions:. Application in a Step-by-Step. Schleinkofer....549–54

Ceramic Hydrogen Thyratrons, Factors Leading to the Development of. Price, Coolidge....76–80

Ceramics, On the Anomaly in Residual Polarization of $BaTiO_3$. Saito, Yamanaka....70–6

Challenge of Universal Mobile Communications, The. Culbertson....973–9

Characteristic Impedance and Phase Velocity of a Shielded Helical Transmission Line, The. Kirschbaum....444–50

Characteristics Affected by Control-Circuit Resistance, Graphical Analysis of Full-Wave Magnetic Amplifier Control. Murakami, Kikuchi....526–30

Characteristics of Carbon-Film Resistors after 12,000 Hours of Operation, Life. Braner, Easterday....201–07

Charged Magnetic Pulse Modulator, D-C. Manteuffel, Cooper....843–50

Charging of Fine Particles, A Theoretical Analysis of the Effects of an Electric Field on the. Murphy, Adler, Penney....318–26

Circuit, Analog Division. McMurray....606–12

Circuit, Analysis and Design of a Transistor Linear-Delay. Nanavati....577–80

Circuit as Applied to a Conference Repeater, A New Dynamic Impedance Matching. Grandstaff....751–5

Circuit-Field Problems, The Use of Iterated Laplace Transformations in the Solution of Combined. Mulligan....506–11

Circuit for Transformers in Which Nonlinear Effects Are Present, An Equivalent. Lord.580–5; disc. 586

Circuits—A Classification of Half-Wave and Full-Wave Nonreversible and Reversible Self-Saturating Circuits, Magnetic Amplifier. McMurtrie....739–51

Circuits, A Comparison of Sequential and Iterative. McCluskey....1039–44

Circuits for Aircraft Electric Systems, Transistor-Magnetic Control. Pratt....643–50

Circuits, Shunt-Coupled Magnetic-Amplifier. Hubbard....124–31

Circuits, The Numerical Graphical Method in the Design of Multiterminal Switching. Scheinman....515–19

Circular Polarization Duplexer for Millimeter Waves, A. Fellers....934–7

City-Wide Personal Signaling at Allentown-Bethlehem, Pa. Kraus....52–5

Classification of Half-Wave and Full-Wave Nonreversible and Reversible Self-Saturating Circuits, Magnetic Amplifier Circuits—A. McMurtrie....739–51

Closed Circuit Educational Television, A Transmission Design for. Wall....340–6

Vol.
Pt. 1
1959

Closed-Circuit Television in Education, A Design for Using. Almstead........................439–44
Cocentric Annular-Plate Capacitors, Subarea Determination of the Capacitance of. Reitan, Higgins..1002–05
Code Repeater, A Transistorized Pulse. Partridge..826–30
Code Teletypewriter Transmission, An Error-Detection System for 5-Unit-. Barry, Whitman......916–21
(Committee Report) Proposed Standard Test Codes for Magnetic Amplifiers. (59–177).............453–6
(Committee Report) Recommended Symbols for Magnetic Amplifier Papers—A Report. (59–173).................................519–20; disc. 520
(Committee Report) Thermal Life of Enameled Magnet Wire. (58–1205)........................91–102
Common Network Theorems, The General Form of Some. Adamson, El-Sobki.......991–5; disc. 995
Communication and Radar Systems, Marginal Utility and a Criterion of Performance for. Schwartz..117–19
(Communication Switching Systems) A Comparison of Sequential and Iterative Circuits. McCluskey..1039–44
(Communication Switching Systems) A New Dynamic Impedance Matching Circuit as Applied to a Conference Repeater. Grandstaff..............751–5
(Communication Switching Systems) A Small High-Speed Transistor and Ferrite-Core Memory. Shafer, Toy, Priebe........................763–9
(Communication Switching Systems) Considerations Pertaining to the Design of an Electronic Telephone Switching System. Brightman, Tubinis....777–86
(Communication Switching Systems) Evaluation of Solderless Wrapped Connections for Central Office Use. Elliott...........................185–94
(Communication Switching Systems) Global Public Telephone Service—1958. Donald, Chandler..811–17
(Communication Switching Systems) Indialing to PBX Extensions: Application in a Step-by-Step Central Office Area. Schleinkofer...............549–54
(Communication Switching Systems) SPUD, A Stored-Program Universal Demonstrator for Computer Training. Raspanti...................586–94
(Communication Switching Systems) Subdivision of Telephone Traffic. Molnar...............28–45
(Communication Switching Systems) The Numerical-Graphical Method in the Design of Multiterminal Switching Circuits. Scheinman...........515–19
(Communication Theory) An Experimental Study of Detection in Nonstationary Noise. Williams, Thomas..............................678–82
(Communication Theory) Asynchronous Multiplexing. Taylor..............................1054–62
(Communication Theory) Effects of Nonlinearity on Propagation in Ionized Media. Cohn.......942–9
(Communication Theory) Network Response to Transient Frequency Modulation Inputs. Bucher..1017–22
(Communication Theory) Transmission Aspects of an Electronic Switchboard Employing Time-Division Multiplexing. Perkins...................949–53
Communications, A Nonsynchronous System for Mobile Record. Stewart.......................141–7
Communications, Considerations in SSB and ISB Systems for Long-Distance Radiotelegraph. Lyons..921–4
Communications, The Challenge of Universal Mobile. Culbertson..........................973–9
Comparison of Mouth-to-Mouth and Manual Artificial Respiration Techniques, A. Gordon, Frye...86–91
Comparison of Sequential and Iterative Circuits, A. McCluskey..........................1039–44
Compensation of Saturating Sampled-Data Systems, The Stability and. Mullin.......270–7; disc. 277
Complex Switching Networks Incorporating Time Delays, Use of High-Speed Digital Computers to Study Performance of. Chang, George.....982–7
Computation of Correlation and Spectral Functions by Orthogonal Filtering, The. Gilbert........954–9
Computational and Experimental Techniques in Nuclear Reactor Design. Witzig, Stuart, Herwig...155–63
Computer Design of Pulse Transformers, Digital. Lego, Greene, Banic..........................217–20
Computer Procedure for the Design of Optimum Systems, A Generalized. Brown:
  Part I..285–9
  Part II..289–93
Computer, Simulation of Data-Switching Systems on a Digital. Gross.......................796–800
Computer Training, SPUD, A Stored-Program Universal Demonstrator for. Raspanti........586–94
Computers, System Synthesis with the Aid of Digital. Dennis, Nease, Saunders..................512–15
Computers to Study Performance of Complex Switching Networks Incorporating Time Delays, Use of High-Speed Digital. Chang, George.............982–7
(Computing Devices) Application of Finite Integral Transforms to Analog Simulations. Karplus, Stephens..................................152–5
(Computing Devices) Matrix Algebra of Sequential Logic—Matrix Logic III. Schubert.......1074–9
(Computing Devices) Matrix Synthesis of High-Speed Logic. Schubert..........................4–8
(Computing Devices) Ripple-Type Time-Delay Networks Using Elliptic Functions. Kiseda, Ford..996–1002
(Computing Devices) Status of Sampled-Data Systems. Jury..769–77
(Computing Devices) Simultaneous Logical Equations—Matrix Logic IV. Schubert..............1080–3
(Computing Devices) Symmetric Switching Functions—Matrix Logic V. Schubert..............1083–7
(Computing Devices) The Stability and Compensation of Saturating Sampled-Data Systems. Mullin..270–7; disc. 277
(Computing Devices) The Use of Automatic Programing Techniques for Solving Engineering Problems. Carleton, Chackan, Martin.....596–601; disc. 601
Concepts of Multidimensional Space Filters, Basic. Kron..554–61
Conducting Sphere in Alternating Magnetic Fields. Poritsky..................................937–42
Conference Repeater, A New Dynamic Impedance Matching Circuit as Applied to a. Grandstaff..751–5
Configuration and Performance, Relation Between Permanent Magnet. Cioffi......979–81; disc. 981
Configuration, Current Distribution in the Cylindrical Source Plane-Electrode. Comstock, Williams..252–6
Connections for Central Office Use, Evaluation of Solderless Wrapped. Elliott.............185–94
Connections 30 Years After Woodson, Polyphase Meter. Warburton.................398–412; disc. 412
Considerations in SSB and ISB Systems for Long Distance Radiotelegraph Communications. Lyons..921–4
Considerations Pertaining to the Design of an Electronic Telephone Switching System. Brightman, Tubinis..777–86
Constant-Voltage Reset Test, Graphical Evaluation of Magnetic Amplifier Performance Based on. Nitzan..691–7; disc. 697
Control Characteristics Affected by Control-Circuit Resistance, Graphical Analysis of Full-Wave Magnetic Amplifier. Murakami, Kikuchi......526–30
Control Source, Observation of Transients in the Series-Connected Saturable Reactor With High-Impedance. Goldstein..........................521–6
Controller, A Magnetic Amplifier Flow. Darling.373–9
Conversion, Magnetic Amplifier Binary-to-Analog. Danylchuk, Katz..........................909–12
Converter for Current and Voltage Measurements, A Wide Range Volt-Ampere. Hermach, Williams..384–8
Coolant Instrumentation for Pressurized Water Reactor Power Plants. Lihota, Hoppa.............131–6
Core, A Mathematical Model for a Magnetic-Amplifier Reactor. Woodson.............872–7; disc. 877
Core Materials, Effects of Ultrahigh Temperature on Magnetic Properties of. Pasnak, Lundsten..1033–8; disc. 1038
Corner-Reflector Antenna, The Design of the. Neff, Tillman..293–5
Corona Loss from Voltage-Charge Diagrams, Rapid Determination of. Dixon................207–09
Corona on Polyethylene, A Study of the Effects of. McMahon, Maloney, Perkins.............654–62
Correlation and Spectral Functions by Orthogonal Filtering, The Computation of. Gilbert.....954–9
Correlation of Measured and Calculated Substation Ground Grid Resistance. Kinyon..698–701; disc. 701
Coupled Magnetic-Amplifier Circuits, Shunt-. Hubbard..124–31
Coupled Magnetic Amplifiers, Capacitively. Collins..707–12
Coupling Problems in Multistage Magnetic Amplifiers, Elimination of. McMurray..............858–63
Criteria and Tests for Realizability of the Inductance Matrix. Tokad, Reed....................924–6
Current and Voltage Measurements, A Wide Range Volt-Ampere Converter for. Hermach, Williams..384–8
Current Distribution in the Cylindrical Source Plane-Electrode Configuration. Comstock, Williams..252–6
Cylindrical Source Plane-Electrode Configuration, Current Distribution in the. Comstock, Williams..252–6

D

(Data Communication) A High-Volume High-Speed Weather Information Distribution System. Schwenzfeger..............................722–8
(Data Communication) An Error-Detection System for 5-Unit-Code Teletypewriter Transmission. Barry, Whitman..............................916–21
(Data Communication) Detection of Transmission Errors in 5-Level Punched Tape. Steeneck, Caley..1005–09
Data, Measurements of Narrow-Band Noise on Telephone Facilities in Connection with Analog. Edson, Froehlich, Townley........................83–6
Data-Switching Systems on a Digital Computer, Simulation of. Gross.......................796–800
Data Systems, Status of Sampled-. Jury.......769–77
Data Transmission, Application of Telegraph Techniques in. Boggs, Boughtwood............336–40
D-C Charged Magnetic Pulse Modulator. Manteuffel, Cooper..................................843–50
D-C Overpotential Testing as a Maintenance Tool in the Industrial Plant, The Use of. Weddendorf..729–36
D-C Potentials, Precision Integrator for. Pattee..450–3
Delayed Neutron Detection Methods Applied to the Detection and Location of Reactor Fuel Element Failures. Dewes, Childs.................1065–9
Delays From a Single Magnetic Storage Core, Long Time. Hardies........................457–61
Demonstrator for Computer Training, SPUD, A Stored-Program Universal. Raspanti............586–94
Design and Manufacture of Direct Burial Wire, The. Robb, Roberts..................662–6; disc. 666
Design Features of a New Frequency Shift Carrier Telegraph System. Stephens.............927–34
Design for Using Closed-Circuit Television in Education, A. Almstead........................439–44
Design of a Vertical Magnetic Recording Head with Large-Scale Models. Smith..............357–62
Design of an Electronic Telephone Switching System, Considerations Pertaining to the. Brightman, Tubinis..................................777–86
Design of Optimum Systems, A Generalized Computer Procedure for the. Brown:
  Part I..285–9
  Part II..289–93
Design of Polyethylene-Insulated Multipair Telephone Cable. Windeler........................736–9
Design of Pulse Transformers, Digital Computer. Lego, Greene, Banic........................217–20
Design of Random-Noise Transformers. Vore...59–63
Design of the Corner-Reflector Antenna, The. Neff, Tillman..................................293–5
Design Parameters for Optimizing the Efficiency of Thermoelectric Generators Utilizing P-Type and N-Type Lead Telluride. Fritts...........817–20
Designs for Large Organic Reactor Power Plants. Bosworth, Parkins........................964–73
Detection in Nonstationary Noise, An Experimental Study of. Williams, Thomas.............678–82
Detection Methods Applied to the Detection and Location of Reactor Fuel Element Failures, Delayed Neutron. Dewes, Childs.................1065–9
Detection of Transmission Errors in 5-Level Punched Tape. Steeneck, Caley.................1005–09
Detection System for 5-Unit-Code Teletypewriter Transmission, An Error-. Barry, Whitman......916–21
Deterioration of Kraft Pulps Using a Mass Spectrometer, A Study of Thermal. Saito, Hino........602–06
Determination of Corona Loss From Voltage-Charge Diagrams, Rapid. Dixon.................207–09
Determination of the Capacitance of Cocentric Annular-Plate Capacitors, Subarea. Reitan, Higgins..1002–05
Development in Western Europe and in Underdeveloped Areas, A Review of Telephone. Wallenstein..15–21; disc. 21
Development of Ceramic Hydrogen Thyratrons, Factors Leading to the. Price, Coolidge............76–80
Diagrams, Rapid Determination of Corona Loss From Voltage-Charge. Dixon.................207–09
Dielectric Gas, Properties of Octafluorocyclobutane, a. Blodgett..63–6
Dielectrics, Effects of Gamma Radiation at 25 C on Silicone. Curris........................297–308
Digital Computer Design of Pulse Transformers. Lego, Greene, Banic..........................217–20
Digital Computer, Simulation of Data-Switching Systems on a. Gross......................796–800
Digital Computers, System Synthesis with the Aid of. Dennis, Nease, Saunders.................512–15
Digital Computers to Study Performance of Complex Switching Networks Incorporating Time Delays, Use of High-Speed. Chang, George........982–7
Dilemmas in the Magnetic Field, Some Multihorned. Avcin................................1087–95
Diodes, Analysis of Magnetic Amplifiers With. Johannessen..............................485–504
Diodes, Analysis of Magnetic Amplifiers Without. Johannessen..............................471–85
Direct Matrix Inversion Method, A New. Shipley, Coleman....................568–70; disc. 570
Display, Remotely Controlled Electroluminescent Totaling. Lyman, Jones, Leger...............248–52
Distribution in the Cylindrical Source Plane-Electrode Configuration, Current. Comstock, Williams..252–6
Distribution on Iron Losses, The Effect of Flux. Shindler..............................1069–74
Dividers, A Method for Calibration of Precision Voltage. Pinckney..........................182–5
Division Circuit, Analog. McMurray.........606–12

Domestic Meters for Test, Statistical Approaches to Selecting. Dwon, Morris.......413–22; disc. 422
Duplexer for Millimeter Waves, A Circular Polarization. Fellers..........934–7
Dynamic Impedance Matching Circuit as Applied to a Conference Repeater, A New. Grandstaff..751–5

E

Eddy-Current Losses in Solid and Laminated Iron. Agarwal..........169–79; disc. 179
Education, A Design for Using Closed-Circuit Television in. Almstead..........439–44
Educational Television, A Talk-Back System for. Norton, Davidson..........209–14
Educational Television, A Transmission Design for Closed Circuit. Wall..........340–6
Effect of Flux Distribution on Iron Losses, The. Shindler..........1069–74
Effects of Gamma Radiation at 25 C on Silicone Dielectrics. Currin..........297–308
Effects of Operation of Germanium Alloy Junction Transistors Above Rated Conditions. Spradlin..........346–52
Effects of Ultrahigh Temperature on Magnetic Properties of Core Materials. Pasnak, Lundsten..........1033–8; disc. 1038
Efficiency in Fast-Response Magnetic Amplifiers, Volt-Second Transfer. Part II. Pula, Lynn, Ringelman..........8–11
Efficiency of Thermoelectric Generators Utilizing P-Type and N-Type Lead Telluride, Design Parameters for Optimizing the. Fritts.......817–20
83B1 Teletypewriter Selective-Calling System, The. Smith, Votaw, Whitman..........119–24
Electric Analog of Heat Flow in Power Transistors, An. Reese, Grannemann, Durant..........640–3
Electric Field on the Charging of Fine Particles, A Theoretical Analysis of the Effects of an. Murphy, Adler, Penney..........318–26
Electric Systems, Transistor-Magnetic Control Circuits for Aircraft. Pratt..........643–50
(Electrical Insulation) A Study of the Effects of Corona on Polyethylene. McMahon, Maloney, Perkins..........654–62
(Electrical Insulation) A Study of Thermal Deterioration of Kraft Pulps Using a Mass Spectrometer. Saito, Hino..........602–06
(Electrical Insulation) Effects of Gamma Radiation at 25 C on Silicone Dielectrics. Currin......297–308
(Electrical Insulation) Magnet Wire for 600 C Temperature Transformers. Wareham..........66–70
(Electrical Insulation) On the Anomaly in Residual Polarization of $BaTiO_3$ Ceramics. Saito, Yamanaka..........70–6
(Electrical Insulation) Properties of Octafluorocyclobutane, a Dielectric Gas. Blodgett..........63–6
(Electrical Insulation) The Use of D-C Overpotential Testing as a Maintenance Tool in the Industrial Plant. Weddendorf..........729–36
(Electrical Insulation) Thermal Life of Enameled Magnet Wire. (Committee Report)..........91–102
Electrical Insulation, Three Decades of Progress in. Berberich..........237–48
Electrical Protection for Transistorized Equipment. Bodle, Hays..........232–7
Electrode Configuration, Current Distribution in the Cylindrical Source Plane-. Comstock, Williams..........252–6
Electrodynamometer for the Precise Measurement of Voltage, Current, Power, and Energy, The Inductronic. Estoppey..........393–8
Electroluminescent Totaling Display, Remotely Controlled. Lyman, Jones, Leger..........248–52
Electromechanical Impedance: Analogs and Duality. Swift..........81–3
Electronic Network Synthesis of Linear Algebraic Matrix Equations. Horn, Honnell..........1028–32
Electronic Switchboard Employing Time-Division Multiplexing, Transmission Aspects of an. Perkins..........949–53
Electronic Telephone Switching System, Considerations Pertaining to the Design of an. Brightman, Tubinis..........777–86
(Electronics) A Theoretical Analysis of the Effects of an Electric Field on the Charging of Fine Particles. Murphy, Adler, Penney..........318–26
(Electronics) Analysis and Design of a Transistor Linear-Delay Circuit. Nanavati..........577–80
(Electronics) Design of a Vertical Magnetic Recording Head with Large-Scale Models. Smith....357–62
(Electronics) Design of Random-Noise Transformers. Vore..........59–63
(Electronics) Effects of Operation of Germanium Alloy Junction Transistors Above Rated Conditions. Spradlin..........346–52
(Electronics) Factors Leading to the Development of Ceramic Hydrogen Thyratrons. Price, Coolidge..........76–80
(Electronics) Field-Strength Measurements in Parallel-Plate Precipitators. Lagarias..........427–33
(Electronics) High-Voltage Rectifier Transformer Problems. Wilson..........352–7
(Electronics) Life Characteristics of Carbon-Film Resistors after 12,000 Hours of Operation. Braner, Easterday..........201–07
(Electronics) Quality Assurance Program. Hulnick, Harding, Rowinski..........326–35
(Electronics) Relation Between Permanent Magnet Configuration and Performance. Cioffi..........979–81; disc. 981
Element Failures, Delayed Neutron Detection Methods Applied to the Detection and Location of Reactor Fuel. Dewes, Childs..........1065–9
Elimination of Coupling Problems in Multistage Magnetic Amplifiers. McMurray..........858–63
Elliptic Functions, Ripple-Type Time-Delay Networks Using. Kiseda, Ford..........996–1002
Enamel Magnet Wire, The Evaluation of. Saums, Pendleton..........1–4
Enameled Magnet Wire, Thermal Life of. (Committee Report)..........91–102
Engineering Aspects of TASI. Bullington, Fraser..........256–60
Engineering of Personal Radio Signaling Systems, Systems. Strack..........55–9
Engineering Problems, The Use of Automatic Programing Techniques for Solving. Carleton, Chackan, Martin..........596–601; disc. 601
Engineers, A New Approach to Training Telephone. Burnett, Adams..........572–6; disc. 576
Engineers, Industry Schools Its. Waldner..........147–52; disc. 152
Equations, Analysis of Series-Connected Saturable Reactor with Capacitive Loading and Finite Control Resistance by Use of Difference. Bourne, Salihi..........461–71
Equations, Electronic Network Synthesis of Linear Algebraic Matrix. Horn, Honnell.......1028–32
Equations—Matrix Logic IV, Simultaneous Logical. Schubert..........1080–3
Equipment, Electrical Protection for Transistorized. Bodle, Hays..........232–7
Equipment, Pagemaster Receiver and Modulation. Young..........45–52
Equivalent Circuit for Transformers in Which Nonlinear Effects Are Present, An. Lord.....580–5; disc. 586
Error-Detection System for 5-Unit-Code Teletypewriter Transmission, An. Barry, Whitman.......916–21
Errors in 5-Level Punched Tape, Detection of Transmission. Steeneck, Caley..........1005–09
Europe and in Underdeveloped Areas, A Review of Telephone Development in Western. Wallenstein..........15–21; disc. 21
Evaluation of Enamel Magnet Wire, The. Saums, Pendleton..........1–4
Evaluation of Magnetic Amplifier Performance Based on Constant-Voltage Reset Test, Graphical. Nitzan..........691–7; disc. 697
Evaluation of Solderless Wrapped Connections for Central Office Use. Elliott..........185–94
Exchange Cable, Manufacture of Composite Cable Sheath for Telephone. Hughes..........650–4
Experience with Long-Distance Television Fields Used for Retransmission. Braun..........594–6
Experimental Study of Detection in Nonstationary Noise, An. Williams, Thomas..........678–82
Experimental Techniques in Nuclear Reactor Design, Computational and. Witzig, Stuart, Herwig..........155–63
Extensions: Application in a Step-by-Step Central Office Area, Indialing to PBX. Schleinkofer..........549–54

F

Factors Leading to the Development of Ceramic Hydrogen Thyratrons. Price, Coolidge..........76–80
Fast-Response Magnetic Amplifiers, Volt-Second Transfer Efficiency in. Part II. Pula, Lynn, Ringelman..........8–11
Feedback in Magnetic Amplifiers, On the. Finzi, Suozzi:
  Part II. Combined Magnetic and Electric Feedbacks..........136–41
Ferrite-Core Memory, A Small High-Speed Transistor and. Shafer, Toy, Priebe..........763–9
Ferroresonance, Theory of. Salihi...755–62; disc. 762
Field on the Charging of Fine Particles, A Theoretical Analysis of the Effects of an Electric. Murphy, Adler, Penney..........318–26
Field-Strength Measurements in Parallel-Plate Precipitators. Lagarias..........427–33
Fields, Conducting Sphere in Alternating Magnetic. Poritsky..........937–42
Film Resistors after 12,000 Hours of Operation, Life Characteristics of Carbon-. Braner, Easterday..........201–07
Filtering, The Computation of Correlation and Spectral Functions by Orthogonal. Gilbert..........954–9
Filters, Basic Concepts of Multidimensional Space. Kron..........554–61
Fine Particles, A Theoretical Analysis of the Effects of an Electric Field on the Charging of. Murphy, Adler, Penney..........318–26
Finite Control Resistance by Use of Difference Equations, Analysis of Series-Connected Saturable Reactor with Capacitive Loading and. Bourne, Salihi..........461–71
Finite Integral Transforms to Analog Simulations, Application of. Karplus, Stephens..........152–5
5-Level Punched Tape, Detection of Transmission Errors in. Steeneck, Caley..........1005–09
Five-Unit-Code Teletypewriter Transmission, An Error-Detection System for. Barry, Whitman....916–21
Flagstaff-Phoenix TJ Radio System, The. Hidy.821–6
Flow Controller, A Magnetic Amplifier. Darling.373–9
Fluoroscopic Image Intensifiers, Standards for Measurement of Brightness Intensification in. Lusby..........504–05; disc. 506
Flux Distribution on Iron Losses, The Effect of. Schindler..........1069–74
4,000, 6,000, and 11,000 Megacycles, Tests Conducted Over Highly Reflective Terrain at. Oxehufwud..........265–70
Four-Wire Y, A Single-Stator Meter for Two Phases of a. Schwartz..........228–31; disc. 231
Frequencies, Transmission Properties of Polyethylene-Insulated Telephone Cables at Voice and Carrier. Eager, Jachimowicz, Kolodny, Robinson..........618–39; disc. 639
Frequency Modulation Inputs, Network Response to Transient. Bucher..........1017–22
Frequency Power Generator Using Transistors, Ultrasonic. Wei..........1062–5
Frequency Shift Carrier Telegraph System, Design Features of a New. Stephens..........927–34
Fuel Element Failures, Delayed Neutron Detection Methods Applied to the Detection and Location of Reactor. Dewes, Childs..........1065–9
Full-Wave Magnetic Amplifier Control Characteristics Affected by Control-Circuit Resistance, Graphical Analysis of. Murakami, Kikuchi..........526–30
Full-Wave Nonreversible and Reversible Self-Saturating Circuits, Magnetic Amplifier Circuits—A Classification of Half-Wave and. McMurtrie.....739–51
Functional Life Tests of Apparatus as Compared with Insulation Material Tests. Manning..........1107–10; disc. 1110
Functions by Orthogonal Filtering, The Computation of Correlation and Spectral. Gilbert..........954–9
Functions—Matrix Logic V, Symmetric Switching. Schubert..........1083–7
Functions, Ripple-Type Time-Delay Networks Using Elliptic. Kiseda, Ford..........996–1002

G

Gamma Radiation at 25 C on Silicone Dielectrics, Effects of. Currin..........297–308
Gas, Properties of Octafluorocyclobutane, a Dielectric. Blodgett..........63–6
General Form of Some Common Network Theorems, The. Adamson, El-Sobki........991–5; disc. 995
Generalized Computer Procedure for the Design of Optimum Systems, A. Brown:
  Part I..........285–9
  Part II..........289–93
Generator, Transistorized Multifrequency Ringing. Kostelich, Howald..........987–91
Generator Using Transistors, Ultrasonic Frequency Power. Wei..........1062–5
Generators Utilizing P-Type and N-Type Lead Telluride, Design Parameters for Optimizing the Efficiency of Thermoelectric. Fritts..........817–20
Germanium Alloy Junction Transistors Above Rated Conditions, Effects of Operation of. Spradlin..........346–52
Global Public Telephone Service—1958. Donald, Chandler..........811–17
Graphical Analysis of Full-Wave Magnetic Amplifier Control Characteristics Affected by Control-Circuit Resistance. Murakami, Kikuchi..........526–30
Graphical Evaluation of Magnetic Amplifier Performance Based on Constant-Voltage Reset Test. Nitzan..........691–7; disc. 697
Graphical Method in the Design of Multiterminal Switching Circuits, The Numerical-. Scheinman..........515–19
Grid Resistance, Correlation of Measured and Calculated Substation Ground. Kinyon...698–701; disc. 701
Ground Grid Resistance, Correlation of Measured and Calculated Substation. Kinyon.698–701; disc. 701

H

Half-Wave and Full-Wave Nonreversible and Reversible Self-Saturating Circuits, Magnetic Amplifier Circuits—A Classification of. McMurtrie.....739–51
Head with Large-Scale Models, Design of a Vertical Magnetic Recording. Smith..........357–62

Vol.
Pt. 1
195

Heart, A-C Shocks of Varying Parameters Affecting the. Kouwenhoven, Knickerbocker, Chesnut, Milnor, Sass.........163–7; disc. 167
Heat Flow in Power Transistors, An Electric Analog of. Reese, Grannemann, Durant.........640–3
Helical Transmission Line, The Characteristic Impedance and Phase Velocity of a Shielded. Kirschbaum.........444–50
High-Capacity Microwave Relay System, A New. Arnold, Isaac, Mathwich, Privett, Thompson.........712–22
High-Frequency Magnetic Amplifiers for Magnetic Memory Sensing, Miniature. Briggs, Hu..863–72
High-Speed Digital Computers to Study Performance of Complex Switching Networks Incorporating Time Delays, Use of. Chang, George.........982–7
High-Speed Logic, Matrix Synthesis of. Schubert..4–8
High-Speed Transistor and Ferrite-Core Memory, A Small. Shafter, Toy, Priebe.........763–9
High-Voltage Rectifier Transformer Problems. Wilson.........352–7
High-Volume High-Speed Weather Information Distribution System, A. Schwenzfeger.........722–8
Highly Reflective Terrain at 4,000, 6,000, and 11,000 Megacycles, Tests Conducted Over. Oxehufwud.........265–70
Hydrogen Thyratrons, Factors Leading to the Development of Ceramic. Price, Coolidge.........76–80

I

Image Intensifiers, Standards for Measurement of Brightness Intensification in Fluoroscopic. Lusby.........504–05; disc. 506
Impedance: Analogs and Duality, Electromechanical. Swift.........81–3
Impedance and Phase Velocity of a Shielded Helical Transmission Line, The Characteristic. Kirschbaum.........444–50
Impedance Bridge for Surface Temperature Measurement, An. Mouly.........388–92; disc. 392
Impedance Matching Circuit as Applied to a Conference Repeater, A New Dynamic. Grandstaff....751–5
Impedance Telephone Repeater, A New Transistorized Negative-. Dimmer, Roback.........673–8
Impedance Voice Repeater Employing Transistors, A Miniaturized Negative-. Howell.........11–15
Indialing to PBX Extensions: Application in a Step-by-Step Central Office Area. Schleinkofer....549–54
Indication System for Pressurized Reactors, A Rod-Position. Floyd, Reuther.........614–18
Inductance Matrix, Criteria and Tests for Realizability of the. Tokad, Reed.........924–6
Inductance of A-C Magnets from Simple Models. Douglas, Voith.........562–8; disc. 568
Inductronic Electrodynamometer for the Precise Measurement of Voltage, Current, Power, and Energy, The. Estoppey.........393–8
Industrial Plant, The Use of D-C Overpotential Testing as a Maintenance Tool in the. Weddendorf.729–36
Industry Schools Its Engineers. Waldner.........147–52; disc. 152
Inputs, Network Response to Transient Frequency Modulation. Bucher.........1017–22
Instrumentation for Pressurized Water Reactor Power Plants, Coolant. Lihota, Hoppa.........131–6
(Instruments) A Method for Calibration of Precision Voltage Dividers. Pinckney.........182–5
(Instruments) Polyphase Meter Connections 30 Years After Woodson. Warburton...398–412; disc. 412
(Instruments) Precision Integrator for D-C Potentials. Pattee.........450–3
(Instruments) Statistical Approaches to Selecting Domestic Meters for Test. Dwon, Morris.........413–22; disc. 422
Instruments, Taut-Band Suspensions for 250-Degree. Thomander, MacIndoe.........379–83; disc. 383
Insulated Multipair Telephone Cable, Design of Polyethylene-. Windeler.........736–9
Insulated Telephone Cables at Voice and Carrier Frequencies, Transmission Properties of Polyethylene-. Eager, Jachimowicz, Kolodny, Robinson.........618–39; disc. 639
Insulating Materials, The Present Status and Anticipated Progress in the Field of. Callinan.........22–7; disc. 27
Insulation-Aging Calculations, Mathematics of. Whitman, Whitman.........308–16; disc. 316
Insulation Material Tests, Functional Life Tests of Apparatus as Compared with. Manning.........1107–10; disc. 1110
Insulation, Three Decades of Progress in Electrical. Berberich.........237–48
Integral Transforms to Analog Simulations, Application of Finite. Karplus, Stephens.........152–5
Integrator for D-C Potentials, Precision. Pattee..450–3
Intensification in Fluoroscopic Image Intensifiers, Standards for Measurement of Brightness. Lusby.........504–05; disc. 506
Inversion Method, A New Direct Matrix. Shipley, Coleman.........568–70; disc. 570
Inverter, A More Stable 3-Phase Transistor-Core Power. Jewett, Schmidt.........686–91
Ionized Media, Effects of Nonlinearity on Propagation in. Cohn.........942–9
Iron, Eddy-Current Losses in Solid and Laminated. Agarwal.........169–79; disc. 179
Iron Losses, The Effect of Flux Distribution on. Schindler.........1069–74
ISB Systems for Long-Distance Radiotelegraph Communications, Considerations in SSB and. Lyons.........921–4
Iterative Circuits, A Comparison of Sequential and. McCluskey.........1039–44

J

Junction Transistors Above Rated Conditions, Effects of Operation of Germanium Alloy. Spradlin.........346–52

K

K24A Syncroplex Telephone Carrier System, The. Coetsee, Curtis, Halina.........1044–54; disc. 1054
Kraft Pulps Using a Mass Spectrometer, A Study of Thermal Deterioration of. Saito, Hino....602–06

L

Laminated Iron, Eddy-Current Losses in Solid and. Agarwal.........169–79; disc. 179
Laplace Transformations in the Solution of Combined Circuit-Field Problems, The Use of Iterated. Mulligan.........506–11
Lead Telluride, Design Parameters for Optimizing the Efficiency of Thermoelectric Generators Utilizing P-Type and N-Type. Fritts.........817–20
Life Characteristics of Carbon-Film Resistors after 12,000 Hours of Operation. Braner, Easterday.........201–07
Life of Enameled Magnet Wire, Thermal. (Committee Report).........91–102
Life Tests of Apparatus as Compared with Insulation Material Tests, Functional. Manning.........1107–10; disc. 1110
Linear Algebraic Matrix Equations, Electronic Network Synthesis of. Horn, Honnell.........1028–32
Linear-Delay Circuit, Analysis and Design of a Transistor. Nanavati.........577–80
Linear Nonlinearity, Spectra Output of Piecewise. Smith.........543–9
Logic—Matrix Logic III, Matrix Algebra of Sequential. Schubert.........1074–9
Logic, Matrix Synthesis of High-Speed. Schubert..4–8
Logical Equations—Matrix Logic IV, Simultaneous. Schubert.........1080–3
Long-Distance Radiotelegraph Communications, Considerations in SSB and ISB Systems for. Lyons.........921–4
Long-Distance Television Fields Used for Retransmission, Experience with. Braun.........594–6
Long Time Delays from a Single Magnetic Storage Core. Hardies.........457–61
Loss From Voltage-Charge Diagrams, Rapid Determination of Corona. Dixon.........207–09
Losses in Solid and Laminated Iron, Eddy-Current. Agarwal.........169–79; disc. 179

M

Magnet Wire for 600 C Temperature Transformers. Wareham.........66–70
Magnet Wire, The Evaluation of Enamel. Saums, Pendleton.........1–4
Magnet Wire, Thermal Life of Enameled. (Committee Report).........91–102
Magnetic Amplifier Binary-to-Analog Conversion. Danylchuk, Katz.........909–12
Magnetic Amplifier Circuits—A Classification of Half-Wave and Full-Wave Nonreversible and Reversible Self-Saturating Circuits. McMurtrie.........739–51
Magnetic-Amplifier Circuits, Shunt-Coupled. Hubbard.........124–31
Magnetic Amplifier Control Characteristics Affected by Control-Circuit Resistance, Graphical Analysis of Full-Wave. Murakami, Kikuchi.........526–30
Magnetic Amplifier Flow Controller, A. Darling.373–9
Magnetic Amplifier Papers—A Report, Recommended Symbols for. (Committee Report).........519–20; disc. 520
Magnetic Amplifier Performance Based on Constant-Voltage Reset Test, Graphical Evaluation of. Nitzan.........691–7; disc. 697
Magnetic-Amplifier Reactor Core, A Mathematical Model for a. Woodson.........872–7; disc. 877
Magnetic-Amplifier–Silicon-Transistor Power Supply for Missile Application, A. Mokrytzki, Stuart.........835–43
Magnetic Amplifier, Theory of 3-Phase Bridge. Storm, Flairty:
  Part I.........878–95
(Magnetic Amplifiers) A More Stable 3-Phase Transistor-Core Power Inverter. Jewett, Schmidt.686–91
(Magnetic Amplifiers) An Equivalent Circuit for Transformers in Which Nonlinear Effects Are Present. Lord.........580–5; disc. 586
(Magnetic Amplifiers) Analog Division Circuit. McMurray.........606–12
(Magnetic Amplifiers) Analysis of Series-Connected Saturable Reactor with Capacitive Loading and Finite Control Resistance by Use of Difference Equations. Bourne, Salihi.........461–71
Magnetic Amplifiers, Automatic Regulators with Self-Balancing. Geyger.........433–8
Magnetic Amplifiers, Capacitively Coupled. Collins.........707–12
Magnetic Amplifiers, Elimination of Coupling Problems in Multistage. McMurray.........858–63
Magnetic Amplifiers for Magnetic Memory Sensing, Miniature High-Frequency. Briggs, Hu...863–72
(Magnetic Amplifiers) Observation of Transients in the Series-Connected Saturable Reactor With High-Impedance Control Source. Goldstein.....521–6
Magnetic Amplifiers, On the Feedback in. Finzi, Suozzi:
  Part II. Combined Magnetic and Electric Feedbacks.........136–41
Magnetic Amplifiers, Proposed Standard Test Codes for. (Committee Report).........453–6
Magnetic Amplifiers, The Winding Capacitances in. Johansen.........702–07
(Magnetic Amplifiers) Theory of Ferroresonance. Salihi.........755–62; disc. 762
(Magnetic Amplifiers) Two-Step Switching of Saturable Reactors. Lovell.........895–900
Magnetic Amplifiers Using a Resistive Reactor Model, Analysis of. Frankenthal.........850–7; disc. 857
Magnetic Amplifiers, Volt-Second Transfer Efficiency in Fast-Response. Pula, Lynn, Ringelman.....8–11
Magnetic Amplifiers With Diodes, Analysis of. Johannessen.........485–504
Magnetic Amplifiers Without Diodes, Analysis of. Johannessen.........471–85
Magnetic Configuration and Performance, Relation Between Permanent. Cioffi.....979–81; disc. 981
Magnetic Control Circuits for Aircraft Electric Systems, Transistor-. Pratt.........643–50
Magnetic-Core Memories, All-Transistor. Goda, Johnston, Markowitz, Rosenberg, Stuart-Williams.........666–73
Magnetic Field, Some Multihorned Dilemmas in the. Avcin.........1087–95
Magnetic Fields, Conducting Sphere in Alternating. Poritsky.........937–42
Magnetic Properties of Core Materials, Effects of Ultrahigh Temperature on. Pasnak, Lundsten.........1033–8; disc. 1038
Magnetic Pulse Modulator, D-C Charged. Manteuffel, Cooper.........843–50
Magnetic Recording Head with Large-Scale Models, Design of a Vertical. Smith.........357–62
Magnetic Storage Core, Long Time Delays from a Single. Hardies.........457–61
Magnetic Transistor-Amplifiers, Self-Regulation in. Hardies, Van Allen.........905–09
Magnetic Wires, The A-C Resistance of Solid. Gohar.........912–16
Magnets from Simple Models, Inductance of A-C. Douglas, Voith.........562–8; disc. 568
Maintenance Tool in the Industrial Plant, The Use of D-C Overpotential Testing as a. Weddendorf.........729–36
Manual Artificial Respiration Techniques, A Comparison of Mouth-to-Mouth and. Gordon, Frye.86–91
Manufacture of Composite Cable Sheath for Telephone Exchange Cable. Hughes.........650–4
Manufacture of Direct Burial Wire, The Design and. Robb, Roberts.........662–6; disc. 666
Marginal Utility and a Criterion of Performance for Communication and Radar Systems. Schwartz.........117–19
Mass Spectrometer, A Study of Thermal Deterioration of Kraft Pulps Using a. Saito, Hino.........602–06
Matching Circuit as Applied to a Conference Repeater, A New Dynamic Impedance. Grandstaff...751–5
Material Tests, Functional Life Tests of Apparatus as Compared with Insulation. Manning.........1107–10; disc. 1110
Materials, Effects of Ultrahigh Temperature on Magnetic Properties of Core. Pasnak, Lundsten.........1033–8; disc. 1038
Materials, The Present Status and Anticipated Progress in the Field of Insulating. Callinan.........22–7; disc. 27
Mathematical Model for a Magnetic-Amplifier Reactor Core, A. Woodson.........872–7; disc. 877
Mathematics of Insulation-Aging Calculations. Whitman, Whitman.........308–16; disc. 316

Matrix Algebra of Sequential Logic—Matrix Logic III. Schubert..........1074–9
Matrix, Criteria and Tests for Realizability of the Inductance. Tokad, Reed..........924–6
Matrix Equations, Electronic Network Synthesis of Linear Algebraic. Horn, Honnell..........1028–32
Matrix Inversion Method, A New Direct. Shipley, Coleman..........568–70; disc. 570
Matrix Logic IV, Simultaneous Logical Equations—Schubert..........1080–3
Matrix Logic V, Symmetric Switching Functions—. Schubert..........1083–7
Matrix Synthesis of High-Speed Logic. Schubert..4–8
Measured and Calculated Substation Ground Grid Resistance, Correlation of. Kinyon..........698–701; disc. 701
Measurement, An Impedance Bridge for Surface Temperature. Mouly..........388–92; disc. 392
Measurement of Voltage, Current, Power, and Energy, The Inductronic Electrodynamometer for the Precise. Estoppey..........393–8
Measurements, A Wide Range Volt-Ampere Converter for Current and Voltage. Hermach, Williams..........384–8
Measurements in Parallel-Plate Precipitators, Field-Strength. Lagarias..........427–33
Measurements of Narrow-Band Noise on Telephone Facilities in Connection with Analog Data. Edson, Froehlich, Townley..........83–6
Media, Effects of Nonlinearity on Propagation in Ionized. Cohn..........942–9
Megacycles, Tests Conducted Over Highly Reflective Terrain at 4,000, 6,000, and 11,000 Megacycles. Oxehufwud..........265–70
Memories, All-Transistor Magnetic-Core. Goda, Johnston, Markowitz, Rosenberg, Stuart-Williams..........666–73
Memory, A Small High-Speed Transistor and Ferrite-Core. Shafter, Toy, Priebe..........763–9
Memory Sensing, Miniature High-Frequency Magnetic Amplifiers for Magnetic. Briggs, Hu..........863–72
Meter Connections 30 Years After Woodson, Polyphase. Warburton..........398–412; disc. 412
Meter for Two Phases of a 4-Wire Y, A Single-Stator. Schwartz..........228–31; disc. 231
Meters for Test, Statistical Approaches to Selecting Domestic. Dwon, Morris..........413–22; disc. 422
Method, A New Direct Matrix Inversion. Shipley, Coleman..........568–70; disc. 570
Method for Calibration of Precision Voltage Dividers, A. Pinckney..........182–5
Methods Applied to the Detection and Location of Reactor Fuel Element Failures, Delayed Neutron Detection. Dewes, Childs..........1065–9
Methods of Obtaining Time of Response, Some Practical. Van Allen, Covert, Wilson..........900–05
Methods of Operation for Private Telegraph Networks, Synchronous. Diamond..........278–83; disc. 283
Microwave Relay System, A New High-Capacity. Arnold, Isaac, Mathwich, Privett, Thompson..........712–22
Microwave Systems on the Pacific Great Eastern Railway. Doble, Raftis..........800–06
Microwaves and Their Use in Power Systems. Bartlett..........1095–1106; disc. 1106
Millimeter Waves, A Circular Polarization Duplexer for. Fellers..........934–7
Miniature High-Frequency Magnetic Amplifiers for Magnetic Memory Sensing. Briggs, Hu...863–72
Miniaturized Negative-Impedance Voice Repeater Employing Transistors, A. Howell..........11–15
Missile Application, A Magnetic Amplifier–Silicon-Transistor Power Supply for. Mokrytzki, Stuart..........835–43
Mixed Transcendental and Polynomial Transforms. Smith..........786–95
Mobile Communications, The Challenge of Universal. Culbertson..........973–9
Mobile Record Communications, A Nonsynchronous System for. Stewart..........141–7
Model, Analysis of Magnetic Amplifiers Using a Resistive Reactor. Frankenthal..........850–7; disc. 857
Model for a Magnetic-Amplifier Reactor Core, A Mathematical. Woodson..........872–7; disc. 877
Models, Inductance of A-C Magnets from Simple. Douglas, Voith..........562–8; disc. 568
Modulation Equipment, Pagemaster Receiver and. Young..........45–52
Modulator, D-C Charged Magnetic Pulse. Manteuffel, Cooper..........843–50
More About Nonarmored Submarine Cable. Lawton..........367–9
Mouth-to-Mouth and Manual Artificial Respiration Techniques, A Comparison of. Gordon, Frye..........86–91
Multichannel Radio Relays and White Noise, CCITT Recommendations for. Parry..........107–17
Multidimensional Space Filters, Basic Concepts of. Kron..........554–61
Multifrequency Ringing Generator, Transistorized. Kostelich, Howald..........987–91
Multihorned Dilemmas in the Magnetic Field, Some. Avcin..........1087–95
Multiplexing, Asynchronous. Taylor..........1054–62
Multiplexing, Transmission Aspects of an Electronic Switchboard Employing Time-Division. Perkins..........949–53
Multistage Magnetic Amplifiers, Elimination of Coupling Problems in. McMurray..........858–63
Multiterminal Switching Circuits, The Numerical-Graphical Method in the Design of. Scheinman..........515–19

N

N-Type Lead Telluride, Design Parameters for Optimizing the Efficiency of Thermoelectric Generators Utilizing P-Type and. Fritts..........817–20
Narrow-Band Noise on Telephone Facilities in Connection with Analog Data, Measurements of. Edson, Froehlich, Townley..........83–6
Negative-Impedance Telephone Repeater, A New Transistorized. Dimmer, Roback..........673–8
Negative-Impedance Voice Repeater Employing Transistors, A Miniaturized. Howell..........11–15
Network Response to Transient Frequency Modulation Inputs. Bucher..........1017–22
Network Synthesis of Linear Algebraic Matrix Equations, Electronic. Horn, Honnell..........1028–32
Network Theorems, The General Form of Some Common. Adamson, El-Sobki..991–5; disc. 995
Networks, Synchronous Methods of Operation for Private Telegraph. Diamond...278–83; disc. 283
Networks, Synthesis of Cable Simulation. DeMonte..........682–6
Networks Using Elliptic Functions, Ripple-Type Time-Delay. Kiseda, Ford..........996–1002
Neutron Detection Methods Applied to the Detection and Location of Reactor Fuel Element Failures, Delayed. Dewes, Childs..........1065–9
New York, Telex in. Easterlin..........195–201
Noise, An Experimental Study of Detection in Nonstationary. Williams, Thomas..........678–82
Noise, CCITT Recommendations for Multichannel Radio Relays and White. Parry..........107–17
Noise on Telephone Facilities in Connection with Analog Data, Measurements of Narrow-Band. Edson, Froehlich, Townley..........83–6
Noise Transformers, Design of Random-. Vore..59–63
Nonarmored Submarine Cable, More About. Lawton..........367–9
Nonlinear Effects Are Present, An Equivalent Circuit for Transformers in Which. Lord..........580–5; disc. 586
Nonlinear Systems, Recurrence Relations in the Solution of a Certain Class of. Wolf.....830–4; disc. 834
Nonlinearity on Propagation in Ionized Media, Effects of. Cohn..........942–9
Nonlinearity, Spectral Output of Piecewise Linear. Smith..........543–9
Nonlinearity, Statistical Spectral Output of Power Law. Smith..........535–43
Nonsynchronous System for Mobile Record Communications, A. Stewart..........141–7
Nuclear Reactor Design, Computational and Experimental Techniques in. Witzig, Stuart, Herwig..........155–63
(Nucleonic and Radiation Instrumentation) A Rod-Position Indication System for Pressurized Reactors. Floyd, Reuther..........614–18
(Nucleonic and Radiation Instrumentation) Coolant Instrumentation for Pressurized Water Reactor Power Plants. Lihota, Hoppa..........131–6
(Nucleonics) Designs for Large Organic Reactor Power Plants. Bosworth, Parkins..........964–73
Numerical-Graphical Method in the Design of Multiterminal Switching Circuits, The. Scheinman..........515–19

O

Observation of Transients in the Series-Connected Saturable Reactor With High-Impedance Control Source. Goldstein..........521–6
Octafluorocyclobutane, a Dielectric Gas, Properties of. Blodgett..........63–6
Office Use, Evaluation of Solderless Wrapped Connections for Central. Elliott..........185–94
156 Kc, A Transposition System for Carrier Systems Up to. Kirkland..........612–14
One-Watt Solar Power Plant, A. Smith..........530–5
Optimum Systems, A Generalized Computer Procedure for the Design of. Brown:
  Part I..........285–9
  Part II..........289–93
Optimum Transmission-Line Pulse Transformer, The. Young, Schatz, Woodford..........220–3
Organic Reactor Power Plants, Designs for Large. Bosworth, Parkins..........964–73
Orthogonal Filtering, The Computation of Correlation and Spectral Functions by. Gilbert..........954–9
Oscillator Scan Stability, Two-. Joseph..........1022–8
Output of Piecewise Linear Nonlinearity, Spectral. Smith..........543–9
Output of Power Law Nonlinearity, Statistical Spectral. Smith..........535–43
Overpotential Testing as a Maintenance Tool in the Industrial Plant, The Use of D-C. Weddendorf..........729–36

P

P-Type and N-Type Lead Telluride, Design Parameters for Optimizing the Efficiency of Thermoelectric Generators Utilizing. Fritts..........817–20
Pacific Great Eastern Railway, Microwave Systems on the. Doble, Raftis..........800–06
Pagemaster Receiver and Modulation Equipment. Young..........45–52
Parallel-Plate Precipitators, Field-Strength Measurements in. Lagarias..........427–33
Parameters Affecting the Heart, A-C Shocks of Varying. Kouwenhoven, Knickerbocker, Chesnut, Milnor, Sass..........163–7; disc. 167
Parameters for Optimizing the Efficiency of Thermoelectric Generators Utilizing P-Type and N-Type Lead Telluride, Design. Fritts..........817–20
Particles, A Theoretical Analysis of the Effects of an Electric Field on the Charging of Fine. Murphy, Adler, Penney..........318–26
PBX Extensions: Application in a Step-by-Step Central Office Area, Indialing to. Schleinkofer....549–43
Performance Based on Constant-Voltage Reset Test, Graphical Evaluation of Magnetic Amplifier. Nitzan..........691–7; disc. 697
Performance for Communication and Radar Systems, Marginal Utility and a Criterion of. Schwartz..........117–19
Permanent Magnet Configuration and Performance, Relation Between. Cioffi..........979–81; disc. 981
Personal Radio Signaling Systems, Systems Engineering of. Strack..........55–9
Personal Signaling at Allentown-Bethlehem, Pa., City-Wide. Kraus..........52–5
Phase Velocity of a Shielded Helical Transmission Line, The Characteristic Impedance and. Kirschbaum..........444–50
Phases of a 4-Wire Y, A Single-Stator Meter for Two. Schwartz..........228–31; disc. 231
Phoenix TJ Radio System, The Flagstaff-. Hidy..........821–6
Plane-Electrode Configuration, Current Distribution in the Cylindrical Source. Comstock, Williams..........252–6
Plate Capacitors, Subarea Determination of the Capacitance of Cocentric Annular-. Reitan, Higgins..........1002–05
Polarization Duplexer for Millimeter Waves, A Circular. Fellers..........934–7
Polarization of $BaTiO_3$ Ceramics, On the Anomaly in Residual. Saito, Yamanaka..........70–6
Polyethylene, A Study of the Effects of Corona on. McMahon, Maloney, Perkins..........654–62
Polyethylene-Insulated Multipair Telephone Cable, Design of. Windeler..........736–9
Polyethylene-Insulated Telephone Cables at Voice and Carrier Frequencies, Transmission Properties of. Eager, Jachimowicz, Kolodny, Robinson..........618–39; disc. 639
Polynomial Transforms, Mixed Transcendental and. Smith..........786–95
Polyphase Meter Connections 30 Years After Woodson. Warburton..........398–412; disc. 412
Potentials, Precision Integrator for D-C. Pattee..450–3
Power and Energy, The Inductronic Electrodynamometer for the Precise Measurement of Voltage, Current. Estoppey..........393–8
Power Generator Using Transistors, Ultrasonic Frequency. Wei..........1062–5
Power Inverter, A More Stable 3-Phase Transistor-Core. Jewett, Schmidt..........686–91
Power Law Nonlinearity, Statistical Spectral Output of. Smith..........535–43
Power Plant, A One-Watt Solar. Smith..........530–5
Power Plants, Coolant Instrumentation for Pressurized Water Reactor. Lihota, Hoppa..........131–6
Power Plants, Designs for Large Organic Reactor. Bosworth, Parkins..........964–73
Power Supply for Missile Application, A Magnetic-Amplifier–Silicon-Transistor. Mokrytzki, Stuart..........835–43
Power Systems, Microwaves and Their Use in. Bartlett..........1095–1106; disc. 1106
Power Transistors, An Electric Analog of Heat Flow in. Reese, Grannemann, Durant..........640–3
Precipitators, Field-Strength Measurements in Parallel-Plate. Lagarias..........427–33
Precision Integrator for D-C Potentials. Pattee..450–3
Precision Voltage Dividers, A Method for Calibration of. Pinckney..........182–5
Present Status and Anticipated Progress in the Field of Insulating Materials, The. Callinan..........22–7; disc. 27

Pressurized Reactors, A Rod-Position Indication System for. Floyd, Reuther.......................614-18
Pressurized Water Reactor Power Plants, Coolant Instrumentation for. Lihota, Hoppa.......131-6
Private Telegraph Networks, Synchronous Methods of Operation for. Diamond.......278-83; disc. 283
Problems in Multistage Magnetic Amplifiers, Elimination of Coupling. McMurray............858-63
Problems, The Use of Iterated Laplace Transformations in the Solution of Combined Circuit-Field. Mulligan...................................506-11
Program, Quality Assurance. Hulnick, Harding, Rowinski...................................326-35
Programing Techniques for Solving Engineering Problems, The Use of Automatic. Carleton, Chackan, Martin.......................596-601; disc. 601
Progress in Electrical Insulation, Three Decades of. Berberich...................................237-48
Progress in the Field of Insulating Materials, The Present Status and Anticipated. Callinan.................................22-7; disc. 27
Propagation in Ionized Media, Effects of Nonlinearity on. Cohn...................................942-9
Properties of Modern Teleprinters with Regard to Signal Transmission. Wusteney...........362-7
Properties of Octafluorocyclobutane, a Dielectric Gas. Blodgett...................................63-6
Proposed Standard Test Codes for Magnetic Amplifiers. (Committee Report).......................453-6
Protection for Transistorized Equipment, Electrical. Bodle, Hays...................................232-7
Public Telephone Service—1958, Global. Donald, Chandler...................................811-17
Pulps Using a Mass Spectrometer, A Study of Thermal Deterioration of Kraft. Saito, Hino.......602-06
Pulse Code Repeater, A Transistorized. Partridge...................................826-30
Pulse Modulator, D-C Charged Magnetic. Manteuffel, Cooper...................................843-50
Pulse Transformer, The Optimum Transmission-Line. Young, Schatz, Woodford.......................220-3
Pulse Transformers, Digital Computer Design of. Lego, Greene, Banic...........................217-20
Punched Tape, Detection of Transmission Errors in 5-Level. Steeneck, Caley................1005-09

Q

Quality Assurance Program. Hulnick, Harding, Rowinski...................................326-35

R

Radar Systems, Marginal Utility and a Criterion of Performance for Communication and. Schwartz...................................117-19
Radiation at 25 C on Silicone Dielectrics, Effect on Gamma. Currin.....................297-308
(Radio Communication) A Circular Polarization Duplexer for Millimeter Waves. Fellers.....934-7
(Radio Communication) A New High-Capacity Microwave Relay System. Arnold, Isaac, Mathwich, Privett, Thompson.......................712-22
(Radio Communication) A Transistorized Pulse Code Repeater. Partridge.....................826-30
(Radio Communication) An Approach to the Study of Automatic Target Tracking. Pode.........806-11
(Radio Communication) City-Wide Personal Signaling at Allentown-Bethlehem, Pa. Kraus........52-5
(Radio Communication) Microwave Systems on the Pacific Great Eastern Railway. Doble, Raftis...................................800-06
(Radio Communication) Pagemaster Receiver and Modulation Equipment. Young............45-52
(Radio Communication) Synchronous Methods of Operation for Private Telegraph Networks. Diamond.....................278-83; disc. 283
(Radio Communication) Tests Conducted Over Highly Reflective Terrain at 4,000, 6,000, and 11,000 Megacycles. Oxehufwud.................265-70
(Radio Communication) The Challenge of Universal Mobile Communications. Culbertson......973-9
(Radio Communication) The Design of the Corner-Reflector Antenna. Neff, Tillman.........293-5
Radio Relays and White Noise, CCITT Recommendations for Multichannel. Parry............107-17
Radio Signaling Systems, Systems Engineering of Personal. Strack.........................55-9
Radio System, The Flagstaff-Phoenix TJ. Hidy...................................821-6
Radiotelegraph Communications, Considerations in SSB and ISB Systems for Long-Distance. Lyons...................................921-4
Railway, Microwave Systems on the Pacific Great Eastern. Doble, Raftis..................800-06
Random-Noise Transformers, Design of. Vore...................................59-63
Rapid Determination of Corona Loss From Voltage-Charge Diagrams. Dixon..................207-09
Reactor Core, A Mathematical Model for a Magnetic-Amplifier. Woodson...........872-7; disc. 877
Reactor Design, Computational and Experimental Techniques in Nuclear. Witzig, Stuart, Herwig...................................155-63
Reactor Fuel Element Failures, Delayed Neutron Detection Methods Applied to the Detection and Location of. Dewes, Childs...............1065-9
Reactor Model, Analysis of Magnetic Amplifiers Using a Resistive. Frankenthal..........850-7; disc. 857
Reactor Power Plants, Coolant Instrumentation for Pressurized Water. Lihota, Hoppa.........131-6
Reactor Power Plants, Designs for Large Organic. Bosworth, Parkins.......................964-73
Reactors, A Rod-Position Indication System for Pressurized. Floyd, Reuther...............614-18
Reactors, Two-Step Switching of Saturable. Lovell...................................895-900
Realizability of the Inductance Matrix, Criteria and Tests for. Tokad, Reed...............924-6
Receiver and Modulation Equipment, Pagemaster. Young...................................45-52
Recommendations for Multichannel Radio Relays and White Noise, CCITT. Parry...........107-17
Recommended Symbols for Magnetic Amplifier Papers—A Report. (Committee Report)...................................519-20; disc. 520
Recording Head with Large-Scale Models, Design of a Vertical Magnetic. Smith..............357-62
Rectifier, A Silicon-Controlled. Bisson, Dyer:
  Part I. Characteristics and Ratings.........102-06
Rectifier Transformer Problems, High-Voltage. Wilson...................................352-7
Recurrence Relations in the Solution of a Certain Class of Nonlinear Systems. Wolf........830-4; disc. 834
Reflector Antenna, The Design of the Corner-. Neff, Tillman...................................293-5
Regulation in Magnetic Transistor-Amplifiers, Self-. Hardies, Van Allen..................905-09
Regulators with Self-Balancing Magnetic Amplifiers, Automatic. Geyger.....................433-8
Relation Between Permanent Magnet Configuration and Performance. Cioffi...........979-81; disc. 981
Relations in the Solution of a Certain Class of Nonlinear Systems, Recurrence. Wolf...................................830-4; disc. 834
Relay System, A New High-Capacity Microwave. Arnold, Isaac, Mathwich, Privett, Thompson...................................712-22
Relays and White Noise, CCITT Recommendations for Multichannel Radio. Parry.............107-17
Remotely Controlled Electroluminescent Totaling Display. Lyman, Jones, Leger...........248-52
Repeater, A New Dynamic Impedance Matching Circuit as Applied to a Conference. Grandstaff...................................751-5
Repeater, A New Transistorized Negative-Impedance Telephone. Dimmer, Roback..............673-8
Repeater, A Transistorized Pulse Code. Partridge...................................826-30
Repeater Employing Transistors, A Miniaturized Negative-Impedance Voice. Howell........11-15
Reperforator-Teletypewriter TT-195()/FG. Frick...................................369-73
Reset Test, Graphical Evaluation of Magnetic Amplifier Performance Based on Constant-Voltage. Nitzan...................................691-7; disc. 697
Residual Polarization of $BaTiO_3$ Ceramics, On the Anomaly in. Saito, Yamanaka.............70-6
Resistance by Use of Difference Equations, Analysis of Series-Connected Saturable Reactor with Capacitive Loading and Finite Control. Bourne, Salihi...................................461-71
Resistance, Correlation of Measured and Calculated Substation Ground Grid. Kinyon...................................698-701; disc. 701
Resistance of Solid Magnetic Wires, The A-C. Gohar...................................912-16
Resistive Reactor Model, Analysis of Magnetic Amplifiers Using a. Frankenthal......850-7; disc. 857
Resistors after 12,000 Hours of Operation, Life Characteristics of Carbon-Film. Braner, Easterday...................................201-07
Response of Tapered Transmission Lines, The Transient. Young, Schatz, Woodford........223-8; disc. 228
Response, Some Practical Methods of Obtaining Time of. Van Allen, Covert, Wilson...........900-05
Response to Transient Frequency Modulation Inputs, Network. Bucher.....................1017-22
Retransmission, Experience with Long-Distance Television Fields Used for. Braun...........594-6
Reversible Self-Saturating Circuits, Magnetic Amplifier Circuits—A Classification of Half-Wave and Full-Wave Nonreversible and. McMurtrie.....739-51
Review of Telephone Development in Western Europe and in Underdeveloped Areas, A. Wallenstein...................................15-21; disc. 21
Ringing Generator, Transistorized Multifrequency. Kostelich, Howald.......................987-91
Ripple-Type Time-Delay Networks Using Elliptic Functions. Kiseda, Ford................996-1002
Rod-Position Indication System for Pressurized Reactors, A. Floyd, Reuther...............614-18

S

(Safety) A Comparison of Mouth-to-Mouth and Manual Artificial Respiration Techniques. Gordon, Frye...................................86-91
(Safety) A-C Shocks of Varying Parameters Affecting the Heart. Kouwenhoven, Knickerbocker, Chesnut, Milnor, Sass.................163-7; disc. 167
Sampled-Data Systems, Status of. Jury.......769-77
Sampled-Data Systems, The Stability and Compensation of Saturating. Mullin...........270-7; disc. 277
Saturable Reactor with Capacitive Loading and Finite Control Resistance by Use of Difference Equations, Analysis of Series-Connected. Bourne, Salihi...................................461-71
Saturable Reactor with High-Impedance Control Source, Observation of Transients in the Series-Connected. Goldstein...................521-6
Saturable Reactors, Two-Step Switching of. Lovell...................................895-900
Saturating Sampled-Data Systems, The Stability and Compensation of. Mullin........270-7; disc. 277
Scan Stability, Two-Oscillator. Joseph.........1022-8
Schools Its Engineers, Industry. Waldner...................................147-52; disc. 152
Selecting Domestic Meters for Test, Statistical Approaches to. Dwon, Morris.....413-22; disc. 422
Selective-Calling System, The 83B1 Teletypewriter. Smith, Votaw, Whitman..................119-24
Self-Balancing Magnetic Amplifiers, Automatic Regulators with. Geyger.......................433-8
Self-Regulation in Magnetic Transistor-Amplifiers. Hardies, Van Allen......................905-09
Self-Saturating Circuits, Magnetic Amplifier Circuits—A Classification of Half-Wave and Full-Wave Nonreversible and Reversible. McMurtrie...................................739-51
(Semiconductor Metallic Rectifiers) A Silicon-Controlled Rectifier. Bisson, Dyer:
  Part I. Characteristics and Ratings.........102-06
(Semiconductor Metallic Rectifiers) Ultrasonic Frequency Power Generator Using Transistors. Wei...................................1062-5
Sensing, Miniature High-Frequency Magnetic Amplifiers for Magnetic Memory. Briggs, Hu..863-72
Sequential and Iterative Circuits, A Comparison of. McCluskey...................................1039-44
Sequential Logic—Matrix Logic III, Matrix Algebra of. Schubert.............................1074-9
Series-Connected Saturable Reactor with Capacitive Loading and Finite Control Resistance by Use of Difference Equations, Analysis of. Bourne, Salihi...................................461-71
Series-Connected Saturable Reactor with High-Impedance Control Source, Observation of Transients in the. Goldstein...................521-6
Sheath for Telephone Exchange Cable, Manufacture of Composite Cable. Hughes..............650-4
Shift Carrier Telegraph System, Design Features of a Frequency. Stephens....................927-34
Shocks of Varying Parameters Affecting the Heart, A-C. Kouwenhoven, Knickerbocker, Chesnut, Milnor, Sass.......................163-7; disc. 167
Shunt-Coupled Magnetic-Amplifier Circuits. Hubbard...................................124-31
Signal Transmission, Properties of Modern Teleprinters with Regard to. Wusteney............362-7
Signaling at Allentown-Bethlehem, Pa., City-Wide Personal. Kraus...........................52-5
Signaling Systems, Systems Engineering of Personal Radio. Strack...........................55-9
Silicon-Controlled Rectifier, A. Bisson, Dyer:
  Part I. Characteristics and Ratings.........102-06
Silicon Dielectrics, Effects of Gamma Radiation at 25 C on. Currin.......................297-308
Silicon-Transistor Power Supply for Missile Application, A Magnetic-Amplifier-. Mokrytski, Stuart...................................835-43
Simulation Networks, Synthesis of Cable. DeMonte...................................682-6
Simulation of Data-Switching Systems on a Digital Computer. Gross.......................796-800
Simultaneous Logical Equations—Matrix Logic IV. Schubert...............................1080-3
Single-Stator Meter for Two Phases of a 4-Wire Y. Schwartz..................228-31; disc. 231
600 C Temperature Transformers, Magnetic Wire for. Wareham...........................66-70
Small High-Speed Transistor and Ferrite-Core Memory, A. Shafer, Toy, Priebe...............763-9
Solar Power Plant, A One-Watt. Smith........530-5
Solderless Wrapped Connections for Central Office Use, Evaluation of. Elliott..............185-94
Solid and Laminated Iron, Eddy-Current Losses in. Agarwal.................169-79; disc. 179
Solid Magnetic Wires, The A-C Resistance of. Gohar...................................912-16
(Solid State Devices) A One-Watt Solar Power Plant. Smith...................................530-5
(Solid State Devices) An Electric Analog of Heat Flow in Power Transistors. Reese, Grannemann, Durant...................................640-3

(Solid State Devices) Design Parameters for Optimizing the Efficiency of Thermoelectric Generators Utilizing P-Type and N-Type Lead Telluride. Fritts..........817–20
(Solid State Devices) Remotely Controlled Electroluminescent Totaling Display. Lyman, Jones, Leger..........248–52
Solution of a Certain Class of Nonlinear Systems, Recurrence Relations in the. Wolf...830–4; disc. 834
Solution of Combined Circuit-Field Problems, The Use of Iterated Laplace Transformations in the. Mulligan 506–11
Some Multihorned Dilemmas in the Magnetic Field. Avcin..........1087–95
Some Practical Methods of Obtaining Time of Response. Van Allen, Covert, Wilson..........900–05
Source Plane-Electrode Configuration, Current Distribution in the Cylindrical. Comstock, Williams.. 252–6
Space Filters, Basic Concepts of Multidimensional. Kron..........554–61
Spectral Functions by Orthogonal Filtering, The Computation of Correlation and. Gilbert...954–9
Spectral Output of Piecewise Linear Nonlinearity. Smith..........543–9
Spectral Output of Power Law Nonlinearity, Statistical. Smith..........535–43
Spectrometer, A Study of Thermal Deterioration of Kraft Pulps Using a Mass. Saito, Hino...602–06
Sphere in Alternating Magnetic Fields, Conducting. Poritsky..........937–42
SPUD, A Stored-Program Universal Demonstrator for Computer Training. Raspanti..........586–94
SSB and ISB Systems for Long-Distance Radiotelegraph Communications, Considerations in. Lyons 921–4
Stability and Compensation of Saturating Sampled-Data Systems, The. Mullin......270–7; disc. 277
Stability, Two-Oscillator Scan. Joseph.......1022–8
Stable 3-Phase Transistor-Core Power Inverter, A More. Jewett, Schmidt..........686–91
Standard Test Codes for Magnetic Amplifiers, Proposed. (Committee Report)..........453–6
Standards for Measurement of Brightness Intensification in Fluoroscopic Image Intensifiers. Lusby.... ....504–05; disc. 506
Statistical Approaches to Selecting Domestic Meters for Test. Dwon, Morris..........413–22; disc. 422
Statistical Spectral Output of Power Law Nonlinearity. Smith..........535–43
Status and Anticipated Progress in the Field of Insulating Materials, The Present. Callinan........ ....22–7; disc. 27
Status of Sampled-Data Systems. Jury..........769–77
Storage Core, Long Time Delays From a Single Magnetic. Hardies..........457–61
Stored-Program Universal Demonstrator for Computer Training, SPUD, A. Raspanti..........586–94
Study of Automatic Target Tracking, An Approach to the. Pode..........806–11
Study of Detection in Nonstationary Noise, An Experimental. Williams, Thomas..........678–82
Study of the Effects of Corona on Polyethylene, A. McMahon, Maloney, Perkins..........654–62
Study of Thermal Deterioration of Kraft Pulps Using a Mass Spectrometer, A. Saito, Hino......602–06
Subarea Determination of the Capacitance of Cocentric Annular-Plate Capacitors. Reitan, Higgins.... 1002–05
Subdivision of Telephone Traffic. Molnar......28–45
Submarine Cable, More About Nonarmored. Lawton 367–9
Substation Ground Grid Resistance, Correlation of Measured and Calculated. Kinyon.......... ....698–701; disc. 701
Surface Temperature Measurement, An Impedance Bridge for. Mouly..........388–92; disc. 392
Suspensions for 250-Degree Instruments, Taut-Band. Thomander, MacIndoe.........379–83; disc. 383
Switchboard Employing Time-Division Multiplexing, Transmission Aspects of an Electronic. Perkins 949–53
Switching Circuits, The Numerical-Graphical Method in the Design of Multiterminal. Scheinman.. 515–19
Switching Functions—Matrix Logic V, Symmetric. Schubert..........1083–7
Switching Networks Incorporating Time Delays, Use of High-Speed Digital Computers to Study Performance of Complex. Chang, George....... 982–7
Switching of Saturable Reactors, Two-Step. Lovell 895–900
Switching System, A Small Automatic Teletypewriter. Shimmin, Vanderlippe, Whitman.........959–64
Switching System, Considerations Pertaining to the Design of an Electronic Telephone. Brightman, Tubinis..........777–86
Switching Systems on a Digital Computer, Simulation of Data-. Gross..........796–800
Symbols for Magnetic Amplifier Papers—A Report, Recommended. (Committee Report).......... ....519–20; disc. 520
Symmetric Switching Functions—Matrix Logic V. Schubert..........1083–7
Synchronous Methods of Operation for Private Telegraph Networks. Diamond. 278–83; disc. 283
Syncroplex Telephone Carrier System, The K24A. Coetsee, Curtis, Halina.......1044–54; disc. 1054
Synthesis of Cable Simulation Networks. DeMonte 682–6
Synthesis of High-Speed Logic, Matrix. Schubert..4–8
Synthesis of Linear Algebraic Matrix Equations, Electronic Network. Horn, Honnell.....1028–32
Synthesis with the Aid of Digital Computers, System. Dennis, Nease, Saunders..........512–15
System, A High-Volume High-Speed Weather Information Distribution. Schwenzfeger..........722–8
System, A Small Automatic Teletypewriter Switching. Shimmin, Vanderlippe, Whitman.........959–64
System for Carrier Systems Up to 156 Kc, A Transposition. Kirkland..........612–14
System for Mobile Record Communications, A Nonsynchronous. Stewart..........141–7
System Synthesis with the Aid of Digital Computers. Dennis, Nease, Saunders..........512–15
Systems, A Generalized Computer Procedure for the Design of Optimum. Brown:
  Part I..........285–9
  Part II..........289–93
Systems Engineering of Personal Radio Signaling Systems. Strack..........55–9
Systems on a Digital Computer, Simulation of Data-Switching. Gross..........796–800
Systems on the Pacific Great Eastern Railway, Microwave. Doble, Raftis..........800–06
Systems, Recurrence Relations in the Solution of a Certain Class of Nonlinear. Wolf.......... ....830–4; disc. 834
Systems, Status of Sampled-Data. Jury.......769–77
Systems, The Stability and Compensation of Saturating Sampled-Data. Mullin..........270–7; disc. 277

T

Talk-Back System from Educational Television, A. Norton, Davidson..........209–14
Tape, Detection of Transmission Errors in 5-Level Punched. Steeneck, Caley..........1005–09
Tapered Transmission Lines, The Transient Response of. Young, Schatz, Woodford.........223–8; disc. 228
Target Tracking, An Approach to the Study of Automatic. Pode..........806–11
TASI, Engineering Aspects of. Bullington, Fraser 256–60
Taut-Band Suspensions for 250-Degree Instruments. Thomander, MacIndoe.........379–83; disc. 383
Techniques in Nuclear Reactor Design, Computational and Experimental. Witzig, Staurt, Herwig...... 155–63
Telegraph Networks, Synchronous Methods of Operation for Private. Diamond..........278–83; disc. 283
Telegraph System, Application of a New Carrier. Collins, Pitts..........1009-16; disc. 1016
Telegraph System, Design Features of a New Frequency Shift Carrier. Stephens..........927–34
(Telegraph Systems) A Nonsynchronous System for Mobile Record Communications. Stewart..141–7
(Telegraph Systems) A Small Automatic Teletypewriter Switching System. Shimmin, Vanderlippe, Whitman..........959–64
(Telegraph Systems) More About Nonarmored Submarine Cable. Lawton..........367–9
(Telegraph Systems) Properties of Modern Teleprinters with Regard to Signal Transmission. Wusteney 362–7
(Telegraph Systems) Reperforator-Teletypewriter TT-195()/FG. Frick..........369–73
(Telegraph Systems) Telex in New York. Easterlin 194–201
(Telegraph Systems) The 83B1 Teletypewriter Selective-Calling System. Smith, Votaw, Whitman..119–24
Telegraph Techniques in Data Transmission, Application of. Boggs, Boughtwood..........336–40
Telegraph Terminal, A Transistorized 20-Channel Carrier. Grybowski, Vieth..........260–5
Telephone Cable, Design of Polyethylene-Insulated Multipair. Windeler..........736–9
Telephone Cables at Voice and Carrier Frequencies, Transmission Properties of Polyethylene-Insulated. Eager, Jachimowicz, Kolodny, Robinson........ ....618–39; disc. 639
Telephone Carrier System, The K24A Syncroplex. Coetsee, Curtis, Halina.......1044–54; disc. 1054
Telephone Development in Western Europe and in Underdeveloped Areas, A Review of. Wallenstein..........15–21; disc. 21
Telephone Engineers, A New Approach to Training. Burnett, Adams..........572–6; disc. 576
Telephone Exchange Cable, Manufacture of Composite Cable Sheath for. Hughes..........650–4
Telephone Facilities in Connection with Analog Data, Measurements of Narrow-Band Noise on. Edson, Froehlich, Townley..........83–6
Telephone Repeater, A New Transistorized Negative-Impedance, Dimmer, Roback..........673–8
Telephone Service—1958, Global Public. Donald, Chandler..........811–17
Telephone Switching System, Considerations Pertaining to the Design of an Electronic. Brightman, Tubinis..........777–86
Telephone Traffic, Subdivision of. Molnar.....28–45
Teleprinters with Regard to Signal Transmission, Properties of Modern. Wusteney..........362–7
Teletypewriter Selective-Calling System, The 83B1. Smith, Votaw, Whitman..........119–24
Teletypewriter Switching System, A Small Automatic. Shimmin, Vanderlippe, Whitman.........959–64
Teletypewriter Transmission, An Error-Detection System for 5-Unit-Code. Barry, Whitman..916–21
Teletypewriter TT-195()/FG, Reperforator-. Frick 369–73
Television, A Talk-Back System for Educational. Norton, Davidson..........209–14
Television, A Transmission Design for Closed Circuit Educational. Wall..........340–6
Television Fields Used for Retransmission, Experience with Long-Distance. Braun..........594–6
Television in Education, A Design for Using Closed-Circuit. Almstead..........439–44
Telex in New York. Easterlin..........194–201
Telluride, Design Parameters for Optimizing the Efficiency of Thermoelectric Generators Utilizing P-Type and N-Type Lead. Fritts.........817–20
Temperature Measurement, An Impedance Bridge for Surface. Mouly..........388–92; disc. 392
Temperature on Magnetic Properties of Core Materials, Effects of Ultrahigh. Pasnak, Lundsten........ ....1033-8; disc. 1038
Temperature Transformers, Magnet Wire for 600 C. Wareham..........66–70
Terminal, A Transistorized 20-Channel Carrier Telegraph. Grybowski, Vieth..........260–5
Test Codes for Magnetic Amplifiers, Proposed Standard. (Committee Report)..........453–6
Test, Graphical Evaluation of Magnetic Amplifier Performance Based on Constant-Voltage Reset. Nitzan..........691–7; disc. 697
Test, Statistical Approaches to Selecting Domestic Meters for. Dwon, Morris......413–22; disc. 422
Testing as a Maintenance Tool in the Industrial Plant, The Use of D-C Overpotential. Weddendorf 729–36
Tests Conducted Over Highly Reflective Terrain at 4,000, 6,000, and 11,000 Megacycles. Oxehufwud 265–70
Tests for Realizability of the Inductance Matrix, Criteria and. Tokad, Reed..........924–6
Tests of Apparatus as Compared with Insulation Material Tests, Functional Life. Manning....... ....1107–10; disc. 1110
Theorems, The General Form of Some Common Network. Adamson, El-Sobki....991–5; disc. 995
Theoretical Analysis of the Effects of an Electric Field on the Charging of Fine Particles, A. Murphy, Adler, Penney..........318–26
Theory of Ferroresonance. Salihi...755–62; disc. 762
Theory of 3-Phase Bridge Magnetic Amplifier:
  Part I. Storm, Flairty..........878–95
Thermal Deterioration of Kraft Pulps Using a Mass Spectrometer, A Study of. Saito, Hino....602–06
Thermal Life of Enameled Magnet Wire. (Committee Report)..........91–102
Thermoelectric Generators Utilizing P-Type and N-Type Lead Telluride, Design Parameters for Optimizing the Efficiency of. Fritts..........817–20
Three Decades of Progress in Electrical Insulation. Berberich..........237–48
Three-Phase Bridge Magnetic Amplifier, Theory of:
  Part I. Storm, Flarity..........878–95
Three-Phase Transistor-Core Power Inverter, A More Stable. Jewett, Schmidt..........686–91
Thyratrons, Factors Leading to the Development of Ceramic Hydrogen. Price, Coolidge.......76–80
Time-Delay Networks Using Elliptic Functions, Ripple-Type. Kiseda, Ford..........996–1002
Time Delays From a Single Magnetic Storage Core, Long. Hardies..........457–61
Time Delays, Use of High-Speed Digital Computers to Study Performance of Complex Switching Networks Incorporating. Chang, George..........982–7
Time-Division Multiplexing, Transmission Aspects of an Electronic Switchboard Employing. Perkins.. 949–53
Time of Response, Some Practical Methods of Obtaining. Van Allen, Covert, Wilson..........900–05
TJ Radio System, The Flagstaff-Phoenix. Hidy..... 821–6
Totaling Display, Remotely Controlled Electroluminescent. Lyman, Jones, Leger..........248–52
Tracking, An Approach to the Study of Automatic Target. Pode..........806–11
Traffic, Subdivision of Telephone. Molnar......28–45
Training Telephone Engineers, A New Approach to. Burnett, Adams..........572–6; disc. 576
Transcendental and Polynomial Transforms, Mixed. Smith..........786–95

Vol.
Pt. 1
1959

Transfer Efficiency in Fast-Response Magnetic Amplifiers, Volt-Second. Part II. Pula, Lynn, Ringelman........8-11
Transformations in the Solution of Combined Circuit-Field Problems, The Use of Iterated Laplace. Mulligan........506-11
Transformer Problems, High-Voltage Rectifier. Wilson........352-7
Transformer, The Optimum Transmission-Line Pulse. Young, Schatz, Woodford........220-3
Transformers, Design of Random-Noise. Vore..59-63
Transformers, Digital Computer Design of Pulse. Lego, Greene, Banic........217-20
Transformers in Which Nonlinear Effects Are Present, An Equivalent Circuit for. Lord. 580-5; disc. 586
Transformers, Magnet Wire for 600 C Temperature. Wareham........66-70
Transforms, Mixed Transcendental and Polynomial. Smith........786-95
Transforms to Analog Simulations, Application of Finite Integral. Karplus, Stephens........152-5
Transient Frequency Modulation Inputs, Network Response to. Bucher........1017-22
Transient Response of Tapered Transmission Lines, The. Young, Schatz, Woodford........223-8; disc. 228
Transients in the Series-Connected Saturable Reactor With High-Impedance Control Source. Goldstein........521-6
Transistor-Amplifiers, Self-Regulation in Magnetic. Hardies, Van Allen........905-09
Transistor and Ferrite-Core Memory, A Small High-Speed. Shafer, Toy, Priebe........763-9
Transistor-Core Power Inverter, A More Stable 3-Phase. Jewett, Schmidt........686-91
Transistor Linear-Delay Circuit, Analysis and Design of a. Nanavati........577-80
Transistor-Magnetic Control Circuits for Aircraft Electric Systems. Pratt........643-50
Transistor Magnetic-Core Memories, All-. Goda, Johnston, Markowitz, Rosenberg, Staurt-Williams........666-73
Transistor Power Supply for Missile Application, A Magnetic-Amplifier–Silicon-. Mokrytzki, Stuart........835-43
Transistorized Equipment, Electrical Protection for. Bodle, Hays........232-7
Transistorized Multifrequency Ringing Generator. Kostelich, Howald........987-91
Transistorized Negative-Impedance Telephone Repeater, A New. Dimmer, Roback........673-8
Transistorized Pulse Code Repeater, A. Partridge........826-30
Transistorized 20-Channel Carrier Telegraph Terminal, A. Grybowski, Vieth........260-5
Transistors, A Miniaturized Negative-Impedance Voice Repeater Employing. Howell........11-15
Transistors Above Rated Conditions, Effects of Operation of Germanium Alloy Junction. Spradlin........346-52
Transistors, An Electric Analog of Heat Flow in Power. Reese, Grannemann, Durant........640-3
Transistors, Ultrasonic Frequency Power Generator Using. Wei........1062-5
Transmission, An Error-Detection System for 5-Unit-Code Teletypewriter. Barry, Whitman........916-21
Transmission, Application of Telegraph Techniques in Data. Boggs, Boughtwood........336-40
Transmission Aspects of an Electronic Switchboard Employing Time-Division Multiplexing. Perkins........949-53
Transmission Design for Closed Circuit Educational Television, A. Wall........340-6
Transmission Errors in 5-Level Punched Tape, Detection of. Steeneck, Caley........1005-09
ransmission-Line Pulse Transformer, The Optimum. Young, Schatz, Woodford........220-3
Transmission Line, The Characteristic Impedance and and Phase Velocity of a Shielded Helical. Kirschbaum........444-50
Transmission Lines, The Transient Response of Tapered. Young, Schatz, Woodford........223-8; disc. 228
Transmission, Properties of Modern Teleprinters with Regard to Signal. Wusteney........362-7
Transmission Properties of Polyethylene-Insulated Telephone Cables at Voice and Carrier Frequencies. Eager, Jachimowicz, Kolodny, Robinson................618-39; disc. 639
Transposition System for Carrier Systems Up to 156 Kc, A. Kirkland........612-14
12,000 Hours of Operation, Life Characteristics of Carbon-Film Resistors after. Braner, Easterday........201-07
20-Channel Carrier Telegraph Terminal, A Transistorized. Grybowski, Vieth........260-5
25 C on Silicone Dielectrics, Effects of Gamma Radiation at. Currin........297-308
250-Degree Instruments, Taut-Band Suspensions for. Thomander, MacIndoe........379-83; disc. 383
Two-Oscillator Scan Stability. Joseph........1022-8
Two Phases of a 4-Wire Y, A Single-Stator Meter for. Schwartz........228-31; disc. 231
Two-Step Switching of Saturable Reactors. Lovell........895-900

U

Ultrahigh Temperature on Magnetic Properties of Core Materials, Effects of. Pasnak, Lundstern................1033-8; disc. 1038
Ultrasonic Frequency Power Generator Using Transistors. Wei........1062-5
Universal Mobile Communications, The Challenge of. Culbertson........973-9
Use of Automatic Programing Techniques for Solving Engineering Problems, The. Carleton, Chackan, Martin........596-601; disc. 601
Use of D-C Overpotential Testing as a Maintenance Tool in the Industrial Plant, The. Weddendorf. 729-36
Use of High-Speed Digital Computers to Study Performance of Complex Switching Networks Incorporating Time Delays. Chang, George........982-7
Use of Iterated Laplace Transformations in the Solution of Combined Circuit-Field Problems, The. Mulligan........506-11
Utility and a Criterion of Performance for Communication and Radar Systems, Marginal. Schwartz........117-19

V

Velocity of a Shielded Helical Transmission Line, The Characteristic Impedance and Phase. Kirschbaum........444-50
Vertical Magnetic Recording Head with Large-Scale Models, Design of a. Smith........357-62
Voice Repeater Employing Transistors, A Miniaturized Negative-Impedance. Howell........11-15
Volt-Ampere Converter for Current and Voltage Measurements, A Wide Range. Hermach, Williams........384-8
Volt-Second Transfer Efficiency in Fast-Response Magnetic Amplifiers. Part II. Pula, Lynn, Ringelman........8-11
Voltage-Charge Diagrams, Rapid Determination of Corona Loss From. Dixon........207-09
Voltage, Current, Power, and Energy, The Inductronic Electrodynamometer for the Precise Measurement of. Estoppey........393-8
Voltage Dividers, A Method for Calibration of Precision. Pinckney........182-5
Voltage Measurements, A Wide Range Volt-Ampere Converter for Current and. Hermach, Williams........384-8
Voltage Reset Test, Graphical Evaluation of Magnetic Amplifier Performance Based on Constant-. Nitzan........691-7; disc. 697

W

Water Reactor Power Plants, Coolant Instrumentation for Pressurized. Lihota, Hoppa........131-6
Waves, A Circular Polarization Duplexer for Millimeter. Fellers........934-7
Weather Information Distribution System, A High-Volume High-Speed. Schwenzfeger........722-8
Western Europe and in Underdeveloped Areas, A Review of Telepnone Development in. Wallenstein........15-21; disc. 21
White Noise, CCITT Recommendations for Multichannel Radio Relays and. Parry........107-17
Wide Range Volt-Ampere Converter for Current and Voltage Measurements, A. Hermach, Williams........384-8
Winding Capacitances in Magnetic Amplifiers, The. Johansen........702-07
(Wire Communications) A Miniaturized Negative-Impedance Voice Repeater Employing Transistors. Howell........11-15
(Wire Communications) A New Approach to Training Telephone Engineers. Burnett, Adams................572-6; disc. 576
(Wire Communications) A New Transistorized Negative Impedance Telephone Repeater. Dimmer, Roback........673-8
(Wire Communications) A Review of Telephone Development in Western Europe and in Underdeveloped Areas. Wallenstein........15-21; disc. 21
(Wire Communications) A Transposition System for Carrier Systems Up to 156 Kc. Kirkland..612-14
(Wire Communications) Design of Polyethylene-Insulated Multipair Telephone Cable. Windeler. 736-9
(Wire Communications) Electrical Protection for Transistorized Equipment. Bodle, Hays....232-7
(Wire Communications) Engineering Aspects of TASI. Bullington, Fraser........256-60
(Wire Communications) Industry Schools Its Engineers. Waldner........147-52; disc. 152
(Wire Communications) Manufacture of Composite Cable Sheath for Telephone Exchange Cable. Hughes........650-4
(Wire Communications) Synthesis of Cable Simulation Networks. DeMonte........682-6
(Wire Communications) The Design and Manufacture of Direct Burial Wire. Robb, Roberts................662-6; disc. 666
(Wire Communications) The K24A Syncroplex Telephone Carrier System. Coetsee, Curtis, Halina................1044-54; disc. 1054
(Wire Communications) Transistorized Multifrequency Ringing Generator. Kostelich, Howald...987-91
(Wire Communications) Transmission Properties of Polyethylene-Insulated Telephone Cables at Voice Carrier Frequencies. Eager, Jachimowicz, Kolodny, Robinson........618-39; disc. 639
Wire for 600 C Temperature Transformers, Magnet. Wareham........66-70
Wire, The Evaluation of Enamel Magnet. Saums, Pendleton........1-4
Wire, Thermal Life of Enameled Magnet. (Committee Report)........91-102
Wires, The A-C Resistance of Solid Magnetic. Gohar........912-16
Woodson, Polyphase Meter Connections 30 Years After. Warburton........398-412; disc. 412
Wrapped Connection for Central Office Use, Evaluation of Solderless. Elliott........185-94

2. Author Index

Adams, Leonard C.,; W. C. Burnett. A New Approach to Training Telephone Engineers. (58-1152)................572-6
Adamson, Colin; S. M. El-Sobki. The General Form of Some Common Network Theorems. (59-1180)........991-5; disc. 995
Adler, Felix T.; A. T. Murphy, G. W. Penney. A Theoretical Analysis of the Effects of an Electric Field on the Charging of Fine Particles. (59-102)........318-26
Agarwal, Paul D. Eddy-Current Losses in Solid and Laminated Iron. (59-12)........169-79; disc. 181
Alger, P. L. Disc........180, 1110
Almstead, Francis E. A Design for Using Closed-Circuit Television in Education. (59-194)..439-44
Anderson, J. Disc........231
Angst, G. Disc........180
Arnold, C. G.; V. E. Isaac, H. R. Mathwich, R. F. Privett, L. E. Thompson. A New High-Capacity Microwave Relay System. (59-874)......712-22
Atkinson, R. W. Disc........27
Avcin, F. Some Multihorned Dilemmas in the Magnetic Field. (59-1187)........1087-95

B

Banic, J. M., Jr.; P. E. Lego, R. L. Greene. Digital Computer Design of Pulse Transformers. (59-48)........217-20
Barker, R. C. Disc........877
Barnes, Thomas D. Disc........412
Barry, P. H.; A. L. Whitman. An Error-Detection System for 5-Unit-Code Teletypewriter Transmission. (59-1106)........916-21
Barthold, L. O. Disc........228
Bartlett, S. C. Microwaves and Their Use in Power Systems. (59-1099)........1095-1106
Berberich, L. J. Three Decades of Progress in Electrical Insulation (59-75)........237-48
Bernstein, T. Disc........216, 296
Bisson, D. K.; R. F. Dyer. A Silicon-Controlled Rectifier: Part I. Characteristics and Ratings. (58-1248)........102-06
Blodgett, F. W. Properties of Octafluorocyclobutane, a Dielectric Gas. (58-336)........63-6
Bodle, D. W.; J. B. Hays, Jr. Electrical Protection for Transistorized Equipment. (59-71)........232-7
Boggs, A.; J. E. Boughtwood. Application of Telegraph Techniques in Data Transmission. (59-107)................336-40
Bosworth, G. H.; W. E. Parkins. Designs for Large Organic Reactor Power Plants. (59-1132)..964-73
Boughtwood, J. E.; A. Boggs. Application of Telegraph Techniques in Data Transmission. (59-107)................336-40
Bourne, H. C.; J. T. Salihi. Analysis of Series-Connected Saturable Reactor with Capacitive Loading and Finite Control Resistance by Use of Difference Equations. (59-170)........461-71
Braner, H.; J. L. Easterday. Life Characteristics of Carbon-Film Resistors after 12,000 Hours of Operation. (59-36)........201-07
Braun, W. L. Experience with Long-Distance Television Fields Used for Retransmission. (59-674)........594-6
Briggs, G. R.; K. C. Hu. Miniature High-Frequency Magnetic Amplifiers for Magnetic Memory Sensing. (59-1042)........863-72
Brightman, B.; M. P. Tubinis. Considerations Pertaining to the Design of an Electronic Telephone Switching System. (59-182)........777-86

Brown, R. Roderick. A Generalized Computer Procedure for the Design of Optimum Systems:
  Part I. (58–234) 285–9
  Part II. (58–235) 289–93
Bruncke, H. P. Disc. 1110
Bucher, T. T. N. Network Response to Transient Frequency Modulation Inputs. (59–1144) 1017–22
Buller, F. H. Disc. 316
Bullington, K.; J. M. Fraser. Engineering Aspects of TASI. (59–70) 256–60
Burnett, William C.; L. C. Adams. A New Approach to Training Telephone Engineers. (58–1152) 572–6
Buscemi, J. V. Disc. 639
Busch, F. H. Disc. 422

C

Caley, H. F.; R. Steeneck. Detection of Transmission Errors in 5-Level Punched Tape. (59–1147) 1005–09
Callinan, Thomas D. The Present Status and Anticipated Progress in the Field of Insulating Materials. (58–1208) 22–7
Carletor, J. T.; N. Chackan, T. W. Martin. The Use of Automatic Programing Techniques for Solving Engineering Problems. (59–641) 596–601
Carter, Gordon K. Disc. 570
Chackan, N.; J. T. Carleton, T. W. Martin. The Use of Automatic Programing Techniques for Solving Engineering Problems. (59–641) 596–601
Chandler, T. A.; D. D. Donald. Global Public Telephone Service—1958. (59–920) 811–17
Chang, Y. N.; O. M. George. Use of High-Speed Digital Computers to Study Performance of Complex Switching Networks Incorporating Time Delays (59-1125) 982–7
Chesnut, R. W.; W. B. Kouwenhoven, G. G. Knickerbocker, W. R. Milnor, D. J. Sass. A-C Shocks of Varying Parameters Affecting the Heart. (59–14) .... 163–7; disc. 168
Childs, J. C.; R. A. Dewes. Delayed Neutron Detection Methods Applied to the Detection and Location of Reactor Fuel Element Failures. (59–1184) 1065–9
Cioffi, P. P. Relation Between Permanent Magnet Configuration and Performance. (59–1120) .... 979–81; disc. 981
Coetsee, B. G.; G. L. Curtis, J. W. Halina. The K24A Syncroplex Telephone Carrier System. (59–1104) 1044–54
Cohn, George I. Effects of Nonlinearity on Propagation in Ionized Media. (59–1145) 942–9
Coleman, Dorothy; R. B. Shipley. A New Direct Matrix Inversion Method. (59–502) .... 568–70; disc. 570
Collins, H. William. Capacitively Coupled Magnetic Amplifiers. (59–785) 707–12
Collins, T. B., Jr.; J. E. Pitts. Application of a New Carrier Telegraph System. (59–1107) .... 1009–16; disc. 1016
Comstock, W. D.; E. M. Williams. Current Distribution in the Cylindrical Source Plane-Electrode Configuration. (59–46) 252–6
Converti, V. Disc. 570
Coolidge, A. W.; H. N. Price. Factors Leading to the Development of Ceramic Hydrogen Thyratrons. (58–188) 76–80
Cooper, R. E.; E. W. Manteuffel. D-C Charged Magnetic Pulse Modulator. (59–1051) 843–50
Covert, Paul W.; R. L. VanAllen, T. G Wilson. Some Practical Methods of Obtaining Time of Response. (59–1061) 900–05
Covey, A. B. Disc. 1016
Culbertson, Alan F. The Challenge of Universal Mobile Communications. (59–1148) 973–9
Currin, C. G. Effects of Gamma Radiation at 25 C on Silicone Dielectrics. (59–117) 297–308
Curtis, G. L.; B. G. Coetsee, J. W. Halina. The K24A Syncroplex Telephone Carrier System. (59–1104) 1044–54

D

Dalziel, Charles F. Disc. 167
Danylchuk, I.; D. Katz. Magnetic Amplifier Binary-to-Analog Conversion. (59–1044) 909–12
Darling, H. E. A Magnetic Amplifier Flow Controller (59–157) 373–9
Davidson, J. B.; R. C. Norton. A Talk-Back System for Educational Television. (59–229) 209–14
DeMonte, Robert W. Synthesis of Cable Simulation Networks. (59–774) 682–6
Dennard, R. H. Disc. 762
Dennis, J. B.; R. F. Nease, R. M. Saunders. System Synthesis with the Aid of Digital Computers. (59–196) 512–15
Desoer, C. A.; S. Kodama. Disc. 277
Dewes, R. A.; J. C. Childs. Delayed Neutron Detection Methods Applied to the Detection and Location of Reactor Fuel Element Failures. (59–1184) 1065–9
Diamond, B. S. Synchronous Methods of Operation for Private Telegraph Networks. (59–61) .... 278–83; disc. 283
Dimmer, R. P.; E. L. Roback. A New Transistorized Negative-Impedance Telephone Repeater. (59–794) 673–8
Dixon, H. S. Rapid Determination of Corona Loss From Voltage-Charge Diagrams. (59–24) 207–09
Doble, R. N.; J. E. Raftis. Microwave Systems on the Pacific Great Eastern Railway. (59–889) 800–06
Donald, D. D.; T. A. Chandler. Global Public Telephone Service—1958. (59–920) 811–17
Dortort, I. K. Disc. 215, 296
Douglas, J. F. H.; R. J. Voith. Inductance of A-C Magnets from Simple Models. (59–235) .... 562–8; disc. 568
Durant, J. R.; J. Reese, W. W. Grannemann. An Electric Analog of Heat Flow in Power Transistors. (59–766) 640–3
Dwon, Larry; J. A. Morris. Statistical Approaches to Selecting Domestic Meters for Test. (59–167) .... 413–22; disc. 426
Dyer, R. F.; D. K. Bisson. A Silicon-Controlled Rectifier:
  Part I. Characteristics and Ratings. (58–1248) 102–06

E

Eager, G. S., Jr.; L. Jachimowicz, I. Kolodny, D. E. Robinson. Transmission Properties of Polyethylene-Insulated Telephone Cables at Voice and Carrier Frequencies. (59–778) .... 618–39; disc. 639
Easterday, J. L.; H. Braner. Life Characteristics of Carbon-Film Resistors after 12,000 Hours of Operation. (59–36) 201–07
Easterlin, Philip R. Telex in New York. (59–37) 194–201
Edson, J. O.; F. E. Froehlich, R. K. Townley. Measurements of Narrow-Band Noise on Telephone Facilities in Connection with Analog Data. (58–1241) 83–6
El-Sobki, S. M.; C. Adamson The General Form of Some Common Network Theorems. (59–1180) .... 991–5; disc. 995
Elliott, S. J. Evaluation of Solderless Wrapped Connections for Central Office Use. (59–33) 185–94
Estoppey, R. F. The Inductronic Electrodynamometer for the Precise Measurement of Voltage, Current, Power, and Energy. (59–165) 393–8

F

Farmer, G. E. Disc. 1106
Fellers, R. G. A Circular Polarization Duplexer for Millimeter Waves. (59–1085) 934–7
Feth, G. C. Disc. 691, 857
Finzi, L. A.; J. J. Suozzi. On the Feedback in Magnetic Amplifiers:
  Part II. Combined Magnetic and Electric Feedbacks (58–1232) 136–41
Flairty, C. W.; H. F. Storm. Theory of 3-Phase Bridge Magnetic Amplifier:
  Part I. (59–1060) 878–95
Floyd, R. C.; J. F. Reuther. A Rod-Position Indication System for Pressurized Reactors. (58–1319) 614–18
Ford, D. J.; J. R. Kiseda. Ripple-Type Time-Delay Networks Using Elliptic Functions. (59–1193) 996–1002
Frankenthal, Shimshon. Analysis of Magnetic Amplifiers Using a Resistive Reactor Model. (59-1046) 850–7
Fraser, J. M.; K. Bullington. Engineering Aspects of TASI (59–70) 256–60
French, R. X. Disc. 168
Frick, D. F. Reperforator-Teletypewriter TT-195()/FG. (59–152) 369–73
Friedlaender, F. J. Disc. 877
Frisco, L. J. Disc. 27
Fritts, R. W. Design Parameters for Optimizing the Efficiency of Thermoelectric Generators Utilizing P-Type and N-Type Lead Telluride. (59–909) 817–20
Froehlich, F. E.; J. O. Edson, R. K. Townley. Measurements of Narrow-Band Noise on Telephone Facilities in Connection with Analog Data. (58–1241) 83–6
Frye, Charles W.; A. S. Gordon. A Comparison of Mouth-to-Mouth and Manual Artificial Respiration Techniques. (58–1012) 86–91

G

George, O. M.; Y. N. Chang. Use of High-Speed Digital Computers to Study Performance of Complex Switching Networks Incorporating Time Delays. (59–1125) 982–7
Geyger, W. A. Automatic Regulators with Self-Balancing Magnetic Amplifiers. (59–174) 433–8
Gilbert, Elmer G. The Computation of Correlation and Spectral Functions by Orthogonal Filtering. (59–1128) 954–9
Goda, Ben T.; W. R. Johnston, S. Markowitz, M. Rosenberg, R. Stuart-Williams. All-Transistor Magnetic-Core Memories. (59–791) 666–73
Gohar, M. Kamal. The A-C Resistance of Solid Magnetic Wires. (59–1086) 912–16
Goldstein, H. L. Observation of Transients in the Series-Connected Saturable Reactor With High-Impedance Control Source. (59–175) 521–6
Goodrich, R. W.; H. W. Hale. Disc. 571
Gordon, Archer S.; C. W. Frye. A Comparison of Mouth-to-Mouth and Manual Artificial Respiration Techniques. (57–1012) 86–91
Grandstaff, O. D. A New Dynamic Impedance Matching Circuit as Applied to a Conference Repeater. (59–824) 751–5
Grannemann, W. W.; J. Reese, J. R. Durant. An Electric Analog of Heat Flow in Power Transistors. (59–766) 640–3
Greene, R. L.; P. E. Lego, J. M. Banic, Jr. Digital Computer Design of Pulse Transformers. (59–48) 217–20
Gross, Eric T. B. Disc. 995
Gross, F. J. Simulation of Data-Switching Systems on a Digital Computer. (59–890) 796–800
Grybowski, T. M.; W. G. Vieth A Transistorized 20-Channel Carrier Telegraph Terminal. (59–85) 260–5

H

Halacsy, Andrew A. Disc. 317, 1038
Hale, H. W.; R. W. Goodrich. Disc. 571
Halina, J. W.; B. G. Coetsee, G. L. Curtis. The K24A Syncroplex Telephone Carrier System. (59–1104) 1044–54
Hansen, E. F. Disc. 180
Hardies, C. E. Long Time Delays from a Single Magnetic Storage Core. (59–176) 457–61
Hardies, C. E.; R. L. Van Allen. Self-Regulation in Magnetic Transistor-Amplifiers. (59–1059) 905–09
Harding, H. G.; R. A Hulnick, J. T. Rowinski. Quality Assurance Program. (59–144) 326–35
Harnden, J. D., Jr. Disc. 296
Harrison, W. C. Disc. 424
Hays, J. B., Jr.; D. W. Bodle. Electrical Protection for Transistorized Equipment. (59–71) 232–7
Hermach, F. L.; E. S. Williams. A Wide Range Volt-Ampere Converter for Current and Voltage Measurements. (59–161) 384–8
Herwig, L. O.; W. F. Witzig, M. R. Stuart. Computational and Experimental Techniques in Nuclear Reactor Design. (58–1315) 155–63
Hidy, John W. The Flagstaff-Phoenix TJ Radio System. (59–943) 821–6
Higgins, Thomas J. Disc. 568, 834
Higgins, Thomas J.; D. K. Reitan. Subarea Determination of the Capacitance of Cocentric Annular-Plate Capacitors. (59–1186) 1002–05
Hino, T.; Y. Saito. A Study of Thermal Deterioration of Kraft Pulps Using a Mass Spectrometer. (59–116) 602–06
Honnell, Pierre M.; R. E. Horn. Electronic Network Synthesis of Linear Algebraic Matrix Equations. (59–1136) 1028–32
Hoppa, C. B.; J. J. Lihota. Coolant Instrumentation for Pressurized Water Reactor Power Plants. (58–1257) 131–6
Horn, Robert E.; P. M. Honnell. Electronic Network Synthesis of Linear Algebraic Matrix Equations. (59–1136) 1028–32
Howald, B. W.; J. F. Kostelich. Transisturized Multifrequency Ringing Generator. (59–1175) 987–91
Howell, A. S. A Miniaturized Negative-Impedance Voice Repeater Employing Transistors. (57–1097) 11–15
Hu, K. C.; G. R. Briggs. Miniature High-Frequency Magnetic Amplifiers for Magnetic Memory Sensing. (59–1042) 863–72
Hubbard, R. M. Shunt-Coupled Magnetic-Amplifier Circuits. (59–635) 124–31
Hughes, D. A. Manufacture of Composite Cable Sheath for Telephone Exchange Cable. (59–792) 650–4
Hulnick, R. A.; H. G. Harding, J. T. Rowinski. Quality Assurance Program. (59–144) 326–35

Vol.
Pt. 1
195

I

Isaac, V. E.; C. G. Arnold, H. R. Mathwich, R. F. Privett, L. E. Thompson. A New High-Capacity Microwave Relay System. (59–874)......712–22

J

Jachimowicz, L.; G. S. Eager, Jr., I. Kolodny, D. E. Robinson. Transmission Properties of Polyethylene-Insulated Telephone Cables at Voice and Carrier Frequencies. (59–778)..........618–39; disc. 639
Jacobs, J E. Disc..........506
Jewett, W. E.; P. L. Schmidt. A More Stable 3-Phase Transistor-Core Power Inverter. (59–786)..........686–91
Johannessen, Paul R. Analysis of Magnetic Amplifiers Without Diodes. (59–171)..........471–85
Johannessen, Paul R. Analysis of Magnetic Amplifiers With Diodes. (59–172)..........485–504
Johansen, Inge. The Winding Capacitances in Magnetic Amplifiers. (59–776)..........702–07
Johnston, William Robert; B. T. Goda, S. Markowitz, M. Rosenberg, R. Stuart-Williams. All-Transistor Magnetic-Core Memories. (59–791)......666–73
Jones, C. I.; R. C. Lyman, A. Leger. Remotely Controlled Electroluminescent Totaling Display. (59–69)..........248–52
Joseph, Horace M. Two-Oscillator Scan Stability. (59–1179)..........1022–8
Jury, E. I. Status of Sampled-Data Systems. (58–1182)..........769–77

K

Karplus, Walter J.; P. A. Stephens, Jr. Application of Finite Integral Transforms to Analog Simulations. (59–10)..........152–5
Katz, D.; I. Danylchuk. Magnetic Amplifier Binary-to-Analog Conversion. (59–1044)..........909–12
Keever, L. M. Disc..........425
Kikuchi, T.; K. Murakami. Graphical Analysis of Full-Wave Magnetic Amplifier Control Characteristics Affected by Control-Circuit Resistance. (59–169)..........526–30
King, Ellis F. Disc..........147, 576
Kinyon, A. L. Correlation of Measured and Calculated Substation Ground Grid Resistance. (59–805)..........698–701
Kirkland, Bess M. A Transposition System for Carrier Systems Up to 156 Kc. (59–509)..........612–14
Kirschbaum, Herbert S. The Characteristic Impedance and Phase Velocity of a Shielded Helical Transmission Line. (59–266)..........444–50
Kiseda, J. R.; D. J. Ford. Ripple-Type Time-Delay Networks Using Elliptic Functions. (59–1193)..........996–1002
Kitchin, D. W.; P. H. Ware. Disc..........1110
Klein, Marvin M. Disc..........216
Knickerbocker, G. G.; W. B. Kouwenhoven, R. W. Chesnut, W. R. Milnor, D. J. Sass. A-C Shocks of Varying Parameters Affecting the Heart. (59–14)..........163–7; disc. 168
Kodama, S.; C. A. Desoer. Disc..........277
Kolodny, I.; G. S. Eager, Jr., L. Jachimowicz, D. E. Robinson. Transmission Properties of Polyethylene-Insulated Telephone Cables at Voice and Carrier Frequencies. (59–778).....618–39; disc. 639
Kostelich, J. F.; B. W. Howald. Transistorized Multifrequency Ringing Generator. (59–1175)..987–91
Kouwenhoven, W. B.; G. G. Knickerbocker, R. W. Chesnut, W. R. Milnor, D. J. Sass. A-C Shocks of Varying Parameters Affecting the Heart. (59–14)..........163–7; disc. 168
Kraus, C. R. City-Wide Personal Signaling at Allentown-Bethlehem, Pa. (58–813)..........52–5
Kron, Gabriel. Basic Concepts of Multidimensional Space Filters. (59–228)..........554–61
Krzyczkowski, Roman. Disc..........15

L

Lagarias, J. S. Field-Strength Measurements in Parallel-Plate Precipitators. (59–206).....427–33
Lawton, C. S. More About Nonarmored Submarine Cable. (59–153)..........367–9
Lee, Reuben. Disc..........586
Leger, A.; R. C. Lyman, C. I. Jones. Remotely Controlled Electroluminescent Totaling Display. (59–69)..........248–52
Lego, P. E.; R. L. Greene, J. M. Banic, Jr. Digital Computer Design of Pulse Transformers. (59–48)..........217–20
Lihota, J. J.; C. B. Hoppa. Coolant Instrumentation for Pressurized Water Reactor Power Plants. (58–1257)..........131–6
Lord, H. W. An Equivalent Circuit for Transformers in Which Nonlinear Effects Are Present. (59–672)..........580–5; disc. 586
Lovell, B. W. Two-Step Switching of Saturable Reactors. (59–1050)..........895–900
Lucal, C. L. Disc..........423
Lundsten, Richard; M. Pasnak. Effects of Ultrahigh Temperature on Magnetic Properties of Core Materials. (59–1117)..........1033–8; disc. 1039
Lusby, Walter S. Standards for Measurement of Brightness Intensification in Fluoroscopic Image Intensifiers. (57–316)..........504–05; disc. 506
Lyman, R. C.; C. I. Jones, A. Leger. Remotely Controlled Electroluminescent Totaling Display. (59–69)..........248–52
Lynn, G. E.; T. J. Pula, J. F. Ringelman. Volt-Second Transfer Efficiency in Fast-Response Magnetic Amplifiers. Part II. (58–173)..........8–11
Lyon, G. E. Disc..........520
Lyons, Walter. Considerations in SSB and ISB Systems for Long-Distance Radiotelegraph Communications. (59–1091)..........921–4

M

MacIndoe, R. C.; V. S. Thomander. Taut-Band Suspensions for 250-Degree Instruments. (59–159)..........379–83; disc. 384
Maloney, Daniel E.; E. J. McMahon, J. R. Perkins. A Study of the Effects of Corona on Polyethylene. (59–800)..........654–62
Manning, M. L. Functional Life Tests of Apparatus as Compared with Insulation Material Tests. (59–1293)..........1107–10; disc. 1111
Manteuffel, E. W.; R. E. Cooper. D-C Charged Magnetic Pulse Modulator. (59–1051)..........843–50
Marble, W. C. Disc..........571
Markowitz, Seymour; B. T. Goda, W. R. Johnston, M. Rosenberg, R. Sutart-Williams. All-Transistor Magnetic-Core Memories. (59–791)......666–73
Martin, T. W.; J. T. Carleton, N. Chackan. The Use of Automatic Programing Techniques for Solving Engineering Problems. (59–641)..........596–601
Mathwich, H. R.; C. G. Arnold, V. E. Isaac, R. F. Privett, L. E. Thompson. A New High-Capacity Microwave Relay System. (59–874)......712–22
McCluskey, E. J., Jr. A Comparison of Sequential and Iterative Circuits. (59–1160)..........1039–44
McMahon, Eugene J.; D. E. Maloney, J. R. Perkins. A Study of the Effects of Corona on Polyethylene. (59–800)..........654–62
McMurray, William. Analog Division Circuit. (59–668)..........606–12
McMurray, William. Elimination of Coupling Problems in Multistage Magnetic Amplifiers. (59–1053)..........858–63
McMurtrie, David L. Magnetic Amplifier Circuits—A Classification of Half-Wave and Full-Wave Nonreversible and Reversible Self-Saturating Circuits. (59–790)..........739–51
Miller, W. J. Disc..........576
Milnor, W. R.; W. B. Kouwenhoven, G. G. Knickerbocker, R. W. Chesnut, D. J. Sass. A-C Shocks of Varying Parameters Affecting the Heart. (59–14)..........163–7; disc. 168
Mokrytzki, B.; R. A. Stuart. A Magnetic Amplifier–Silicon-Transistor Power Supply for Missile Application. (59–1054)..........835–43
Molnar, Imre. Subdivision of Telephone Traffic. (57–809)..........28–45
Morrill, Weston. Disc..........1038
Morris, J. A.; L. Dwon. Statistical Approaches to Selecting Domestic Meters for Test. (59–167)..........413–22; disc. 426
Mouly, R. J. An Impedance Bridge for Surface Temperature Measurement. (59–158)..........388–92; disc. 393
Mulligan, J. H., Jr. The Use of Iterated Laplace Transformations in the Solution of Combined Circuit-Field Problems. (59–227)..........506–11
Mullin, Francis J. The Stability and Compensation of Saturating Sampled-Data Systems. (59–52)..........270–7; disc. 278
Murakami, K.; T. Kikuchi. Graphical Analysis of Full-Wave Magnetic Amplifier Control Characteristics Affected by Control-Circuit Resistance. (59–169)..........526–30
Murphy, Arthur T.; F. T. Adler, G. W. Penney. A Theoretical Analysis of the Effects of an Electric Field on the Charging of Fine Particles. (59–102)..........318–26

N

Nanavati, R. P. Analysis and Design of a Transistor Linear-Delay Circuit. (59–563)..........577–80
Nease, R. F.; J. B. Dennis, R. M. Saunders. System Synthesis with the Aid of Digital Computers. (59–196)..........512–15
Neff, H. P., Jr.; J. D. Tillman. The Design of the Corner-Reflector Antenna. (59–87)..........293–5
Nitzan, D. Graphical Evaluation of Magnetic Amplifier Performance Based on Constant-Voltage Reset Test. (59–789)..........691–7; disc. 697
Norton, R. C.; J. B. Davidson. A Talk-Back System for Educational Television. (59–229)..........209–14

O

Oxehufwud, A. Tests Conducted Over Highly Reflective Terrain at 4,000, 6,000, and 11,000 Megacycles. (59–86)..........265–70

P

Palmer, G. A. Disc..........425
Parkins, W. E.; G. H. Bosworth. Designs for Large Organic Reactor Power Plants. (59–1132).964–73
Parkins, W. E.; E. F. Weisner. Disc..........215
Parry, Charles A. CCITT Recommendations for Multichannel Radio Relays and White Noise. (59–6)..........107–17
Partridge, Gordon R. A Transistorized Pulse Code Repeater. (59–944)..........826–30
Pasnak, Michael; R. Lundsten. Effects of Ultrahigh Temperature on Magnetic Properties of Core Materials. (59–1117)..........1033–8; disc. 1039
Pattee, John R. Precision Integrator for D-C Potentials. (59–166)..........450–3
Pendleton, W. W.; H. L. Saums. The Evaluation of Enamel Magnet Wire. (58–401)..........1–4
Penney, Gaylord W.; A. T. Murphy, F. T. Adler. A Theoretical Analysis of the Effects of an Electric Field on the Charging of Fine Particles. (59–102)..........318–26
Perkinpine, R. W. Disc..........666
Perkins, J. C., Jr. Transmission Aspects of an Electronic Switchboard Employing Time-Division Multiplexing. (59–210)..........949–53
Perkins, Joseph R.; E. J. McMahon, D. E. Maloney. A Study of the Effects of Corona on Polyethylene. (59–800)..........654–62
Peterson, A. E. Disc..........601
Pinckney, Charles B. A Method for Calibration of Precision Voltage Dividers. (59–22)..........182–5
Pitts, J. E.; T. B. Collins, Jr. Application of New Carrier Telegraph System. (59–1107)..........1009–16; disc. 1016
Pode, Leonard. An Approach to the Study of Automatic Target Tracking. (59–925)..........806–11
Polanecsky, A. J. Disc..........423
Poritsky, H. Conducting Sphere in Alternating Magnetic Fields. (59–1102)..........937–42
Poritsky, H. Disc..........180
Pratt, A. W. Transistor-Magnetic Control Circuits for Aircraft Electric Systems. (59–765)..........643–50
Price, H. N.; A. W. Coolidge. Factors Leading to the Development of Ceramic Hydrogen Thyratrons. (58–188)..........76–80
Priebe, H. F., Jr.; W. L. Shafer, Jr., W. N. Toy. A Small High-Speed Transistor and Ferrite-Core Memory. (59–826)..........763–9
Privett, R. F.; C. G. Arnold, V. E. Isaac, H. R. Mathwich, L. E. Thompson. A New High-Capacity Microwave Relay System. (59–874)......712–22
Pula, T. J.; G. E. Lynn, J. F. Ringelman. Volt-Second Transfer Efficiency in Fast-Response Magnetic Amplifiers. Part II. (58–173)..........8–11

R

Raftis, J. E.; R. N. Doble. Microwave Systems on the Pacific Great Eastern Railway. (59–889)..800–06
Raspanti, M. SPUD, A Stored-Program Universal Demonstrator for Computer Training. (59–665)..........586–94
Reed, Myril B.; Y. Tokad. Criteria and Tests for Realizability of the Inductance Matrix. (59–1090)..........924–6
Reese, James; W. W. Grannemann, J. R. Durant. An Electric Analog of Heat Flow in Power Transistors. (59–766)..........640–3
Reitan, Daniel K.; T. J. Higgins. Subarea Determination of the Capacitance of Cocentric Annular-Plate Capacitors. (59–1186)..........1002–05
Reuther, J. F.; R. C. Floyd. A Rod-Position Indication System for Pressurized Reactors. (58–1319)..........614–18
Ringelman, J. F.; T. J. Pula, G. E. Lynn. Volt-Second Transfer Efficiency in Fast-Response Magnetic Amplifiers. Part II. (58–173)..........8–11
Roback, E. L.; R. P. Dimmer. A New Transistorized Negative-Impedance Telephone Repeater. (59–794)..........673–8
Robb, J. L.; W. L. Roberts. The Design and Manufacture of Direct Burial Wire. (59–777)......662–6

Roberts, W. L.; J. L. Robb. The Design and Manufacture of Direct Burial Wire. (59–777).......662–6
Robinson, D. E.; G. S. Eager, Jr., L. Jachimowicz, I. Kolodny. Transmission Properties of Polyethylene-Insulated Telephone Cables at Voice and Carrier Frequencies. (59–778)..618–39; disc. 639
Rose, Arthur F. Disc..........................1044
Rosenberg, Milton; B. T. Goda, W. R. Johnston, S. Markowitz, R. Stuart-Williams. All-Transistor Magnetic-Core Memories. (59–791)......666–73
Rowinski, J. T.; R. A. Hulnick, H. G. Harding. Quality Assurance Program. (59–144).........326–35

S

Saito, Y.; T. Hino. A Study of Thermal Deterioration of Kraft Pulps Using a Mass Spectrometer. (59–116)..................................602–06
Saito, Y.; S. Yamanaka. On the Anomaly in Residual Polarization of $BaTiO_3$ Ceramics. (58–879).70–6
Salihi, Jalal T. Theory of Ferroresonance. (59–788)............................755–62; disc. 762
Salihi, J. T.; H. C. Bourne. Analysis of Series-Connected Saturable Reactor with Capacitive Loading and Finite Control Resistance by Use of Difference Equations. (59–170)......................461–71
Sass, D. J.; W. B. Kouwenhoven, G. G. Knickerbocker, R. W. Chesnut, W. R. Milnor. A-C Shocks of Varying Parameters Affecting the Heart. (59–14)..............................163–7; disc. 168
Saums, H. L.; W. W. Pendleton. The Evaluation of Enamel Magnet Wire. (58–401).............1–4
Saunders, R. M.; J. B. Dennis, R. F. Nease. System Synthesis with the Aid of Digital Computers. (59–196).................................512–15
Schatz, E. R.; F. J. Young, J. B. Woodford. The Optimum Transmission-Line Pulse Transformer. (59–16).................................220–3
Schatz, E. R.; F. J. Young, J. B. Woodford. The Transient Response of Tapered Transmission Lines. (59–15)...........................223–8; disc. 228
Scheinman, A. H. The Numerical-Graphical Method in the Design of Multiterminal Switching Circuits. (59–1028)..............................515–19
Schindler, M. J. The Effect of Flux Distribution on Iron Losses. (59–1204).................1069–74
Schindler, M. J. Disc............................981
Schleinkofer, G. Norman. Indialing to PBX Extensions: Application in a Step-by-Step Central Office Area. (59–185)................................549–54
Schmidt, P. L.; W. E. Jewett. A More Stable 3-Phase Transistor-Core Power Inverter. (59–786).686–91
Schubert, E. J. Matrix Algebra of Sequential Logic—Matrix Logic III. (59–204)............1074–9
Schubert, E. J. Matrix Synthesis of High-Speed Logic. (58–1205)..................................4–8
Schubert, E. J. Simultaneous Logical Equations—Matrix Logic IV. (59–1192).............1080–3
Schubert, E. J. Symmetric Switching Functions—Matrix Logic V. (59–205)..............1083–7
Schwartz, L. S. Marginal Utility and a Criterion of Performance for Communication and Radar Systems. (58–1264)......................117–19
Schwarz, E. W. A Single-Stator Meter for Two Phases of a 4-Wire Y. (59–17).......228–31; disc. 231
Schwenzfeger, E. E. A High-Volume High-Speed Weather Information Distribution System. (59–812)..................................722–8
Shafer, W. L., Jr.; W. N. Toy, H. F. Priebe, Jr. A Small High-Speed Transistor and Ferrite-Core Memory. (59–826)......................763–9
Shimmin, E. R.; R. A. Vanderlippe, A. L. Whitman. A Small Automatic Teletypewriter Switching System. (59–1105)...............................959–64
Shipley, R. Bruce; D. Coleman. A New Direct Matrix Inversion Method. (59–502)....568–70; disc. 570
Slemon, Gordon S. Disc..........................763
Smith, C. W.; C. J. Votaw, A. L. Whitman. The 83B1 Teletypewriter Selective-Calling System. (58–443)...................................119–24
Smith, D. H. A One-Watt Solar Power Plant. (58–186)...................................530–5
Smith, D. M. Disc............................412
Smith, J. T. Design of a Vertical Magnetic Recording Head with Large-Scale Models. (59–148)..357–62
Smith, Otto J. M. Mixed Transcendental and Polynomial Transforms. (59–917)...........786–95
Smith, Otto J. M. Spectral Output of Piecewise Linear Nonlinearity. (59–264)....................543–9
Smith, Otto J. M. Statistical Spectral Output of Power Law Nonlinearity. (59–265).............535–43
Spradlin, B. C. Effects of Operation of Germanium Alloy Junction Transistors Above Rated Conditions. (59–145)...............................346–52
Steeneck, R.; H. F. Caley. Detection of Transmission Errors in 5-Level Punched Tape. (59–1147)...................................1005–09
Stephens, M. L. Design Features of a New Frequency Shift Carrier Telegraph System. (59–1111)...................................927–34
Stephens, Prewitt A., Jr.; W. J. Karplus. Application of Finite Integral Transforms to Analog Simulations. (59–10)................................152–5
Stevens, R. F. Disc............................701
Stewart, C. H., II. A Nonsynchronous System for Mobile Record Communications. (59–8)...141–7
Storm, H. F. Disc............................521
Storm, H. F.; C. W. Flairty. Theory of 3-Phase Bridge Magnetic Amplifier:
  Part I. (59–1060).........................878–95
Strack, Walter. Systems Engineering of Personal Radio Signaling Systems. (58–989)...............55–9
Stuart, M. R.; W. F. Witzig, L. O. Herwig. Computational and Experimental Techniques in Nuclear Reactor Design. (58–1315)..............155–63
Stuart, R. A.; B. Mokrytzki. A Magnetic Amplifier–Silicon-Transistor Power Supply for Missile Application. (59–1054)........................835–43
Stuart-Williams, Raymond; B. T. Goda, W. R. Johnston, S. Markowitz, M. Rosenberg. All-Transistor Magnetic-Core Memories. (59–791)......666–73
Suozzi, J. J.; L. A. Finzi. On the Feedback in Magnetic Amplifiers:
  Part II. Combined Magnetic and Electric Feedbacks. (58–1232).......................136–41
Swift, Wayne B. Electromechanical Impedance: Analogs and Duality. (58–529)............81–3

T

Taylor, J. E. Asynchronous Multiplexing. (59–1146)...................................1054–62
Templeton, J. G. C. Disc........................423
Thomander, V. S.; R. C. MacIndoe. Taut-Band Suspensions for 250-Degree Instruments. (59–159)..............................379–83; disc. 384
Thomas, John B.; T. R. Williams. An Experimental Study of Detection in Nonstationary Noise. (59–809).................................678–82
Thompson, L. E.; C. G. Arnold, V. E. Isaac, H. R. Mathwich, R. F. Privett. A New High-Capacity Microwave Relay System. (59–874)......712–22
Tillman, J. D.; H. P. Neff, Jr. The Design of the Corner-Reflector Antenna. (59–87)........293–5
Tokad, Yilmaz; M. B. Reed. Criteria and Tests for Realizability of the Inductance Matrix. (59–1090)...................................924–6
Townley, R. K.; J. O. Edson, F. E. Froehlich. Measurements of Narrow-Band Noise on Telephone Facilities in Connection with Analog Data. (58–1241)...................................83–6
Toy, W. N.; W. L. Shafer, Jr., H. F. Priebe, Jr. A Small High-Speed Transistor and Ferrite-Core Memory. (59–826)......................763–9
Tubinis, M. P.; B. Brightman. Considerations Pertaining to the Design of an Electronic Telephone Switching System. (59–182).............777–86

V

Van Allen, Roland L.; P. W. Covert, T. G. Wilson Some Practical Methods of Obtaining Time of Response. (59–1061)...................900–05
Van Allen, R. L.; C. E. Hardies. Self-Regulation in Magnetic Transistor Amplifiers. (59–1059)...................................905–09
Vanderlippe, R. A.; E. R. Shimmin, A. L. Whitman. A Small Automatic Teletypewriter Switching System. (59–1105)...............................959–64
Vieth, W. G.; T. M. Grybowski. A Transistorized 20-Channel Carrier Telegraph Terminal. (59–85)...................................260–5
Voith, R. J.; J. F. H. Douglas. Inductance of A-C Magnets from Simple Models. (59–235)..............................562–8; disc. 568
Vore, M. P. Design of Random-Noise Transformers. (57–1015)................................59–63
Votaw, C. J.; C. W. Smith, A. L. Whitman. The 83B1 Teletypewriter Selective-Calling System. (58–443)...................................119–24

W

Waidelich, Donald L. Disc.......................392
Waldner, C. E. Industry Schools Its Engineers. (58–1235)................................147–52
Walker, H. P. Disc............................383
Wall, W. L. A Transmission Design for Closed Circuit Educational Television. (59–200).........340–6
Wallenstein, G. D. A Review of Telephone Development in Western Europe and in Underdeveloped Areas. (57–786)..................15–21; disc. 21
Warburton, F. W. Polyphase Meter Connections 30 Years After Woodson. (59–168)..............................398–412; disc. 413
Ware, P. H.; D. W. Kitchin. Disc...............1110
Wareham, W. W. Magnet Wire for 600 C Temperature Transformers. (58–117)..................66–70
Weber, Dr. Ernst. Disc.........................179
Weddendorf, W. A. The Use of D-C Overpotential Testing as a Maintenance Tool in the Industrial Plant. (59–511)........................729–36
Wei, Yin-Min. Ultrasonic Frequency Power Generator Using Transistors. (59–1203)............1062–5
Weisner, E. F.; W. E. Parkins. Disc.............215
Weit, James R. Disc...........................728
Werden, A. C., Jr. Disc........................215
Whitman, A. L.; P. H. Barry. An Error-Detection System for 5-Unit-Code Teletypewriter Transmission. (59–1106)........................916–21
Whitman, A. L.; E. R. Shimmin, R. A. Vanderlippe. A Small Automatic Teletypewriter Switching System. (59–1105)...............................959–64
Whitman, A. L.; C. W. Smith, C. J. Votaw. The 83B1 Teletypewriter Selective-Calling System. (58–443)...................................119–24
Whitman, L. C.; W. W. Whitman. Mathematics of Insulation-Aging Calculations. (59–118)..............................308–16; disc. 316
Whitman, W. W.; L. C. Whitman. Mathematics of Insulation-Aging Calculations. (59–118)..............................308–16; disc. 316
Williams, E. M.; W. D. Comstock. Current Distribution in the Cylindrical Source Plane-Electrode Configuration. (59–46)......................252–6
Williams, E. S.; F. L. Hermach. A Wide Range Volt-Ampere Converter for Current and Voltage Measurements. (59–161).......................384–8
Williams, Thomas R.; J. B. Thomas. An Experimental Study of Detection in Nonstationary Noise. (59–809).................................678–82
Wilson, Thomas G.; R. L. VanAllen, P. W. Covert. Some Practical Methods of Obtaining Time of Response. (59–1061).........................900–05
Wilson, T. L. High-Voltage Rectifier Transformer Problems. (59–150).......................352–7
Windeler, A. S. Design of Polyethylene-Insulated Multipair Telephone Cable. (59–793)..........736–9
Wishart, R. S. Disc............................283
Witzig, W. F.; M. R. Stuart, L. O. Herwig. Computational and Experimental Techniques in Nuclear Reactor Design. (58–1315)..............155–63
Wolf, Alfred A. Recurrence Relations in the Solution of a Certain Class of Nonlinear Systems. (59–919)..............................830–4; disc. 834
Woodford, J. B.; F. J. Young, E. R. Schatz. The Optimum Transmission-Line Pulse Transformer. (59–16).................................220–3
Woodford, J. B.; F. J. Young, E. R. Schatz. The Transient Response of Tapered Transmission Lines. (59–15)...........................223–8; disc. 228
Woodson, H. H. A Mathematical Model for a Magnetic-Amplifier Reactor Core. (59–1058)..............................872–7; disc. 878
Wusteney, Herbert. Properties of Modern Teleprinters with Regard to Signal Transmission. (59–154)...................................362–7

Y

Yamanaka, S.; Y. Saito. On the Anomaly in Residual Polarization of $BaTiO_3$ Ceramics. (58–879)...................................70–6
Young, Frederick J. Disc.......................181
Young, F. J.; E. R. Schatz, J. B. Woodford. The Optimum Transmission-Line Pulse Transformer. (59–16).................................220–3
Young, F. J.; E. R. Schatz, J. B. Woodford. The Transient Response of Tapered Transmission Lines. (59–15)...........................223–8; disc. 228
Young, J. W. Pagemaster Receiver and Modulation Equipment. (58–814).....................45–5

Vol.
Pt. 1
1959

The Evaluation of Enamel Magnet Wire

H. L. SAUMS
MEMBER AIEE

W. W. PENDLETON
MEMBER AIEE

Synopsis: This paper is a discussion of the need to broaden the base of test procedures relating to enameled magnet wires. At present, the AIEE Publication no. 57, revised in 1959, establishes a method for determining the relative life of enameled wires as a function of thermal stability.[1] Also available to air evaluation are the AIEE test procedures 65, 510, and 511.[2] These are intended for evaluation of systems of which magnet wire insulations are an integral part. Experience is showing the desirability of providing acceptable test procedures for determining the relative life and the practicability or suitability of magnet wires as functions of other characteristics. Of importance, in this respect, are: 1. compatibility of the wire enamels with other substances, both of insulating function and environmental in nature, which are present in a complete electrical system, 2. thermoplastic flow properties, 3. heat shock characteristics, 4. change of reaction rate of aging properties with increase in temperature, 5. heat softening after extended aging, 6. retention of flexibility after aging, and 7. winding problems reflected in abrasion resistance, flexibility, reaction to rapid elongation, etc.

IT IS NOT the purpose of this paper to discredit or otherwise replace the present AIEE test procedures. They represent excellent beginning tests to evaluate the thermal stability of magnet wire. What is intended is a means for going beyond these tests in a complete evaluation of a wire for practical use at the indicated temperature level. It can be shown, for instance, that some testing done by the suggested methods can lead to erroneous extrapolation of the life-temperature curves unless supporting data are available. It is the purpose of the paper to bring out the important supporting evidence needed to insure the correct thermal rating of a magnet wire enamel. The normal standard tests made by the magnet wire producers for many years are important. However, recent advances in thermal stability of organic materials make important some additional test procedures such as varnish compatibility tests, moisture tests in sealed tubes, effects of other insulating components on wire enamels, and environmental tests. Most of these tests were unnecessary at class 105 C (degrees centigrade) (A) temperatures but are all important at class 130 C (B) and above. The reason for this change of emphasis is that chemical reactions are increased at a rapid rate with increased temperature. For example, the effects of active contaminants can be increased by a factor of 30 or more in raising the temperature from 105 to 155 C. Since aging is often carried out at greatly accelerated temperatures, the possibility of reactions taking place in short periods of time is great. Such "side reactions" can alter the thermal stability picture for high-stability materials to a much greater extent than for standard class 105 C materials.

The effects of incompatibility between magnet wire enamel and varnish are magnified at the higher temperatures. Thus, Formvar can be treated with a number of varnishes, some of which are not compatible with it, but there is no marked lowering of thermal stability while the same varnishes on a polyester enamel can result in a wide variety of life-temperature characteristics. If the postulation is correct that cracks are formed by nonuniform shrinkage of unlike films, the higher temperatures would accentuate this action. Softening or embrittlement of enamel and evolution of active or passive by-products, may also be active in the aging process.

Moisture problems, especially in enclosed systems, become important at the class 130 and class 155 C levels.

The environment problem includes the uses of gases other than air, liquids such as oil, askarel, and others, and solid potting compounds and resins such as epoxy, polyesters, and modifications. The thermal stability of a magnet wire enamel may be greatly affected by the test medium. While it is agreed that air is the most used medium for presently designed apparatus, there is a tendency toward the use of inert and electronegative gases in large transformers, potting compounds, and high temperature liquid insulants for smaller units, and freons and other liquids or gases for hermetically sealed rotating equipment. It is considered important that magnet wire enamels and other insulating components be given (for the benefit of the users) a series of thermal classifications based on environment, type of supporting insulations, and type of electric device. Furthermore, it is necessary to have a list of compatible and noncompatible components to be used or avoided when use is made of the magnet wire enamel being evaluated.

Definition of Magnet Wire

Before attempting to discuss in detail the ramifications of the subject, it seems of value to define magnet wire in order that its properties may be best used and to forestall any misuse derived from misconceptions of its purpose in electric devices.

Magnet wire, as conceived by Joseph Henry more than 125 years ago, is defined as a continuous insulated length of conducting metal capable of being wound either by form or in random manner on a supporting arbor or in or on an insulated ferromagnetic armature of core for the express purpose, in the completed device, of carrying electric current for the establishment of an electric field associated with the magnetic circuit upon which it is wound. The insulation on magnet wire is present only to separate individual turns so that only one wire path is open to the passage of electric current. Magnet wire must have associated insulation to withstand applied voltage to ground or between phases, to cement adjacent turns in a consolidated mass, and to protect against external damage and contamination; see Fig. 1.

Magnet wire insulation is at the major source of heat, the I^2R (current resistance) loss in the conductor. In a given device, it is therefore the most likely insulating component to be subjected to

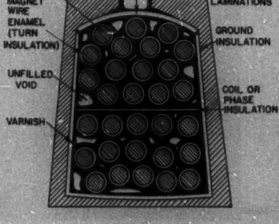


Fig. 1. Function of magnet wire enamel

Paper 58-401, recommended by the AIEE Electrical Insulation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted February 17, 1958; made available for printing September 12, 1958.

H. L. Saums and W. W. Pendleton are with the Anaconda Wire and Cable Company, Muskegon, Mich.

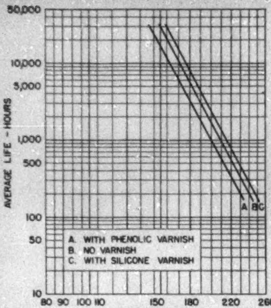


Fig. 2. Effects of varnish type on thermal stability of high-temperature enamel

the "hottest" spot (highest temperature). It is also in a position least able to be cooled and is dependent on associated insulation for conducting its heat to the cooling medium. If the latter is air or a gas, it is important that varnishes or compounds deeply penetrate the windings so that high thermal conductivity materials are present to remove the heat from the wire. If the coolant is a liquid, it is important that all air and moisture be removed before impregnation.

The best advantages of enameled wire in comparison with fiber-insulated types are found in high space factor, excellent flexibility, smooth exterior, high dielectric strength, and often moisture resistance. Even greater space factors are realized when square and rectangular film-insulated wires are used in place of round wire.

The high dielectric strength of magnet wire enamels has tempted some users to eliminate some portions of supporting insulation in the quest for economy of design. This is considered unwise for the following reasons: 1. Magnet wire, unless in very thick grades (triple or quadruple), may have one unavoidable discontinuity of microscopic size in each 10 feet of wire length. 2. Unavoidable winding damage may occur. 3. Expansion and contraction in a finished device in the field may wear the wire in contact with unprotected iron slots. 4. Surge voltages from switching, etc., may exceed the dielectric strength of the wire. 5. Contamination may ground a winding which is separated only by the wire enamel.

The moisture resistance advantage of film insulation over fiber insulation is sometimes of fleeting importance. Some film insulations, possessing moisture resistance at room temperature, are susceptable to attack by moisture under certain high-temperature conditions.

In the magnet wire definition are the words "capable of being wound." It is important that the basic physical and thermal tests be made in any evaluation of thermal stability of an enameled wire. Such properties as flexibility, abrasion resistance, adherence, heat shock, thermoplastic flow, and conductor hardness are important in respect to arriving at specimens which can be properly tested. A serious deficiency in any one property will give test results which are noninterpretable.

Compatibility Problems

At the top of the list of auxiliary data needed beyond the AIEE test data are the problems associated with compatibility. The word "compatibility" has been used to describe the effects of the combination of enamel and varnish and other insulation members when subjected to aging conditions. It has also been used to describe the effects of environmental conditions and presence of contaminants. In other words, compatibility is used to describe how a magnet wire enamel "gets along" with its neighbors, whether they are insulating or contaminating.

The foremost compatibility problem facing the user of the new higher thermal stability enamels is the matching of varnishes or impregnating compounds with the enamels. It is known that varnishes can add to, subtract from, or leave unchanged the thermal stability of a wire enamel; see Fig. 2. If the varnish does not reduce the thermal stability of the enamel it is considered compatible. A complete picture of the thermal stability of a wire enamel would give life-temperature curves for unvarnished specimens; for varnished specimens in at least six categories: silicone, polyester, epoxy, Sylkyd, phenolic, and alkyd; and for motorette or motor tests using the same varnishes and with possible variations in slot insulation and phase insulation. The AIEE twist test has been designed as a screening test, since it is simpler to carry out than the motorette and motor tests.[1] It was envisioned that any varnish incompatibility found in this simple test method would save further testing in the more complex tests. There are, however, incompatibilities which are brought out only by the system tests, and thus it is important that these tests be run in addition to the twist tests for an accurate picture of thermal stability. An important additional test method for compatibility of wire enamels with other insulating components and in the presence of contaminants and environmental components is the sealed tube test; see Fig.

Vol.
Pt. 1
1959

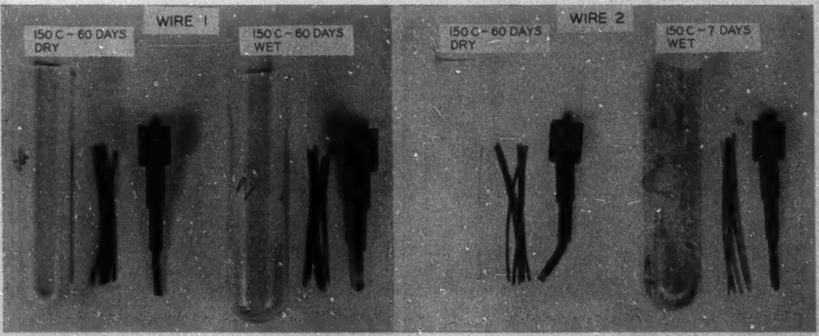
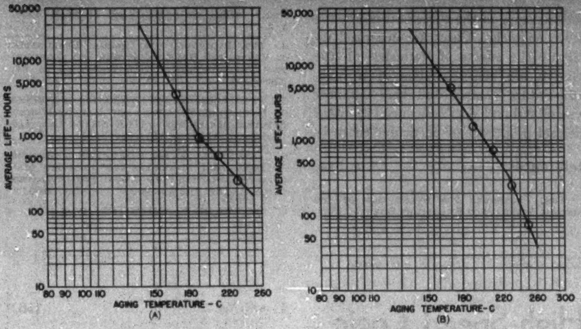


Fig. 3. Sealed tube test method of determining compatibility



A—Decreased reaction rate

B—Increased reaction rate

Fig. 4. Reaction rates at high temperature

3. Wires are wound on a step mandrel whose diameters are one, three, five, and ten times that of the wire. These specimens are removed from the mandrel and placed (either varnished or unvarnished) in glass tubes sealed at the lower end. Into each tube are placed pieces of insulation such as would be used for supporting insulation in an electric machine. The tube is narrowed to a thin neck and its atmosphere adjusted to give dry, moist, or wet conditions in air or in other gases, and it is finally sealed by torch. The tubes are placed in ovens at various temperatures and observed for cracking and discoloration as aging progresses. Any injurious by-products of aging from other components are concentrated in this test and the results represent the worst conditions. They are, however, indicative of possible weaknesses under extreme conditions. Tests of this kind are shown for two different wire types in Fig. 3. Type *1* wire resists the effects of moisture up to 60 days at 150 C. Some discoloration is beginning to show on the straight lengths of wire and on the *1X* mandrel wrap. Type *2*, however, is rapidly attacked so that in 7 days the enamel is partially removed from the wire. Even in the dry condition, where air was used to break a vacuum drawn on the tube prior to sealing the system, type *2* wire shows some hydrolysis effects, presumably from moisture which was not entirely removed by evacuation. Such a wire would be expected to give some difficulties where moisture is not continually being removed from a system by open-air cooling. A similar statement may be made concerning chlorinated supporting insulations where the possibility of HCl as a gas or as an acid may be liberated and concentrated by the lack of free-air cooling. In the latter case, however, even in open apparatus, such members as polyvinyl chloride and neoprene may be objectionable with type *2* enamels. Compatibility of magnet wire enamels with insulating liquids is also a problem of making use of the higher temperature enamels in liquid-filled class 130 C devices. Again, sealed tube tests provide a convenient method for observing the effects of liquid on enamel and enamel on liquid at elevated temperatures and controlled atmospheres.

Effects of Thermal Characteristics of Enamel on Life-Temperature Curves

The assumption is made in the AIEE test procedures that there will be a single reaction rate from the highest testing temperature to the extrapolated use point of the wire. It is experimentally true that this condition holds for many of the materials studied. However, there are some enamels which do not follow a single reaction rate. Following the life-temperature curve from 10,000 hours to lesser periods of time, there may be a point (tentatively identified with the thermoplastic flow point) where there is a mild or a pronounced transition to a second reaction rate which may be greater or less than the longer life reaction rate (at lower temperatures). Figs. 4(A) and 4(B) illustrate these conditions. It is obvious that any reliable extrapolation of these curves to lower temperatures requires three or four points which are well removed from the knee of the curves.

A second phenomenon has been uncovered for still another type of enamel; the apparent discontinuous life-temperature curve, as shown in Fig. 5. At a point near the thermoplastic flow point of the enamel, these characteristics change in a radical manner, so that high-temperature tests extrapolate to a much lower life expectancy than the lower temperature tests. Such an enamel would give excellent service only where no excessive overloads were encountered. These test results all indicate that high temperatures transform the enamel sufficiently to alter its aging rate. It is further indicative of the effects of thermoplastic flow point on the shape of the life-temperature curves. One other possibility is that tests at temperatures exceeding a certain level produce a heat shock condition in the test specimens even though the bending is mild. It is thus important that a wire enamel have sufficient heat shock resistance to allow accelerated aging in the AIEE test procedures.

A recent test program on a material with a rather low thermoplastic flow point, but with a thermal life expectancy of class 180 C levels shows that accelerated aging is not possible due to premature cut-through of the insulation. Such insulation would always be vulnerable to overload heating even though the basic material may be unaffected, chemically, at the time of the overheating.

Summary

That a completely interpretable and useful analysis might be made of a wire enamel, this paper emphasizes the need to broaden the base of testing enameled magnet wire with AIEE test procedures, and the need of compatibility tests involving sealed tubes and other thermal and physical tests on the wire.

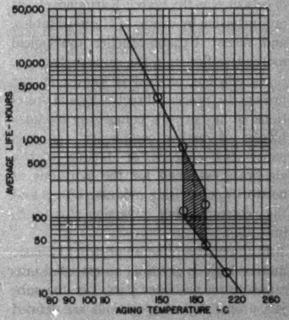


Fig. 5. Discontinuous reaction rate characteristic

It is found that the use of organic enamels at classes 130 C and 155 C necessitates a much closer look at compatibility, thermoplastic flow, heat shock, moisture resistance, and environmental effects than has been given to class 105 C enamels in the past.

The matching of varnish and enamel, the elimination of chemically contaminating supporting insulation, and the proper choice of design and media are important steps in the successful use of the newer enamels. Cases are discussed involving the effects of varnishes on the relative life characteristics of a high-temperature enamel and the nonlinear type of stability data influenced by other thermal properties of particular enamels.

References

1. Test Procedure for Evaluation of the Thermal Stability of Enamelled Wire. "Twist Test," *AIEE Publication no. 57*, Jan. 1959, p. 2.

2. Test Procedure for Thermal Evaluation of Ventilated Dry-Type Power and Distribution Transformers. *AIEE Publication no. 65*, Nov. 1956.

3. Test Procedure for Evaluation of Systems of Insulating Materials for Random-Wound Electric Machinery. *AIEE Standard no. 510* (revision of *AIEE no. 1C*), Nov. 1956.

4. Test Procedure for Evaluation of Systems of Insulating Materials for Electric Machinery Employing Form-Wound Pre-Insulated Coils. *AIEE Publication no. 511*, Oct. 1956.

5. See reference 1, "Switch Test," p. 4.

Matrix Synthesis of High-Speed Logic

E. J. SCHUBERT
ASSOCIATE MEMBER AIEE

Synopsis: The object of this paper is to describe a method for synthesizing function matrices from logical conditions and for detecting redundancy prior to designing the network. Accordingly the paper discusses the following: 1. the definition of function matrices and logical operations; 2. the expansion of logical functions in view of additional logical conditions imposed by another variable; and 3. minimization routines.

CONVENTIONAL Boolean algebra and set-theory as applied for the analysis and minimization of logical networks are limited to a few variables. In electronic equipment with diode and transistor circuitry the number of variables increases rapidly. Boolean equations for these large networks become too cumbersome and the method breaks down eventually due to inefficiency in dealing with the multitude of possible permutations. Graphical representations of sets such as the Veitch-Karnaugh chart, minimization by means of Harvard tables, and recent Urbano-Mueller topological methods suffer from the same restriction.

No efficient method is known at the present to process directly logical functions of a network comprising modular blocks of higher complexity. In this paper the concept of binary matrix algebra is extended for the synthesis of networks comprising modular blocks and the corresponding inversion for minimization.

In the next section an abstract of basic matrix logic is given as an essential prerequisite for the later part. In the subject proper logical operations are defined by a matrix operator $[h]$ with reference to the relation between two variables. By substituting for one of the variables a complete function matrix, a method for network synthesis is derived, called the "expansion theorem." The inverted matrix routine is applicable for the minimization of networks by multiplying the function matrix with a redundancy matrix. The product directly comprises the redundant terms and the irrelevant variables.

Function Matrices

The behavior of a logical network is described explicitly in a truth table by referring to any possible combination of the p independent variables. The set of p variables may assume 2^p states in certain AND or OR conditions called minterm or maxterm, respectively. A minterm N_i is any one of the 2^p distinct AND conditions of all p variables in which these variables are or are not complemented in either of the two possible ways. The suffix of N_i is the octal equivalent of the binary number formed by writing ZERO for each complemented variable and ONE for each uncomplemented variable, in the order of their appearance. Any function Y of p variables can be expressed as a polynomial of all minterms with appropriate coefficients F_i.

$$Y = F_0N_0 + F_1N_1 + \dots + F_qN_q \tag{1A}$$

wherein the coefficients F_i have the values of ONE or ZERO according to the appropriate column for the function Y in the composite truth table.

Using the Einstein convention for the matrices involved, i denotes the dummy index between 0 and $2^p - 1$; z is the fixed index. Capital letters designate the individual terms of the corresponding matrices in index notations.

$$Y_z = F_{zi}N_i \tag{1B}$$

OR Operation. The sum of minterms is by definition an OR operation. Because functions are expressed as sums of minterms, any sum of functions Y_z again will be a sum of minterms.

$$Z = Y_{1i} + Y_{2i} = (F_i + G_i)N_i$$

AND Operation. To obtain an output Z from an AND gate implies that both the Y_1 and Y_2 inputs into the gate are ONE. The state of the output depends on the coefficients of the function matrices, F_i and G_i. Z can be ONE only if the coefficients in both matrices with identical suffix are ONE. All other minterms must be inhibited.

$$Z = Y_{1i}Y_{2i} = (F_iG_i)N_i$$

Negation. This operation implies that each F_{zi} term must be substituted by its complement.

$$F_{zi} + F_{zi}' = 1 \qquad '\dots\text{NOT}$$

$$F_{zi}' = 1 - F_{zi}$$

It is understood that the minterm matrix is 1-dimensional in any case comprising only one row or column, and identical for any operation performed in a particular system. The function matrix may comprise up to 2^{2^p} rows, but in practice only those rows describing utilized outputs are considered.

$$Y_z' = (1 - F_{zi})N_i$$

Certain rules are pertinent to establish the function matrix properly, particularly for those input conditions where not all variables are involved.[1]

Paper 58-1025, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Sacramento, Calif., August 19–22, 1958. Manuscript submitted January 23, 1958; made available for printing September 12, 1958.

E. J. Schubert, formerly with Westinghouse Electric Corporation, Baltimore, Md., is now with the Burroughs Corporation Research Center, Paoli, Pa.

Table I. Logical Operations

Boolean Notation	Matrix Operator [h]				Operation
	H_0	H_1	H_2	H_3	
$Y_0=0$	0	0	0	0	OFF
$Y_1=X_1'X_2'$	1	0	0	0	NOR
$Y_2=X_1X_2$	0	1	0	0	
$Y_3=X_2'$	1	1	0	0	ONLY X_2'
$Y_4=X_1'X_2$	0	0	1	0	
$Y_5=X_1'$	1	0	1	0	ONLY X_1'
$Y_6=X_2'X_1+X_2X_1'$	0	1	1	0	EITHER 3)
$Y_7=X_1'+X_2'$	1	1	1	0	NOT-AND (Sheffer's Stroke)
$Y_{10}=X_1X_2$	0	0	0	1	AND
$Y_{11}=X_1X_2+X_1'X_2$	1	0	0	1	ONLY IF 3)
$Y_{12}=X_1$	0	1	0	1	ONLY X_1
$Y_{13}=X_1+X_2'$	1	1	0	1	
$Y_{14}=X_2$	0	0	1	1	ONLY X_2
$Y_{15}=X_2+X_1'$	1	0	1	1	IF THEN 3)
$Y_{16}=X_1+X_2$	0	1	1	1	OR
$Y_{17}=1$	1	1	1	1	ON

Logical Operations

Functions of two variables are most suitable for representing logical operations.[2] A table of these basic operations may be derived as follows.

Coefficients H_i describe the operation $[h]$ with respect to the individual minterms formed by X_1 and X_2 respectively, according to equations 1(A) and (B), and are listed in Table I. The functions Y_0 OFF and Y_{17}ON are trivial. Four functions, Y_3, Y_5, Y_{12}, and Y_{14}, are truly functions of one of the two variables only. Y_{10} is the basic AND function, Y_{16} is the "inclusive OR" function to be distinguished from Y_6, the "exclusive OR" or "EITHER" function. Several other functions have been defined elsewhere but are not conventional.[3] Recently the NOR function Y_1 has been widely used in RCTL (resistor-coupled-transistor logic) and certain magnetic networks.[4,5]

Completely different is the NOT-AND function Y_7, which has been introduced because of its significance in certain types of redundancy. It is identical with the Sheffer's stroke function.

Expansion Theorem

Expansion of a logical function V of a set of p variables, $U_1 \ldots U_p$, is the process of introducing an additional variable U under the conditions specified by the operator H.

$$V=f(U_1, U_2 \ldots U_{p-1}, U_p) \tag{2A}$$

$$V=F_iN_i \qquad i=0\ldots(2^p-1) \tag{2B}$$

$$W=f(H, V, U) \tag{2C}$$

This resulting output W is a certain function, g, of the enlarged set of $p+1$ variables.

$$W=g(V_1, V_2 \ldots V_p, V_{p+1}) \tag{2D}$$

$$W=G_iN_i+G_jN_j \quad j=2^p\ldots(2^{p+1}-1) \tag{2E}$$

To derive the function matrix $[g]$ directly from $[f]$ and $[h]$, is the concept of the expansion theorem for computing the coefficients $G_{i,j}$ according to the matrix equation

$$\begin{bmatrix} G_i \\ G_j \end{bmatrix} = \begin{bmatrix} H_0 & H_1 \\ H_2 & H_3 \end{bmatrix} \begin{bmatrix} F_i' \\ F_i \end{bmatrix} \tag{3}$$

To prove the relation 3 requires only to develop 2(E). The variables, U and V, may assume four states only.

$$W=H_0U'V'+H_1U'V+H_2UV'+H_3UV \tag{4A}$$

$$W=U'(H_0V'+H_1V)+U(H_2V'+H_3V) \tag{4B}$$

$$W=U'(H_0F_i'+H_1F_i)N_i+U(H_2\times F_i'+H_3F_i)N_i \tag{4C}$$

$$N_i=U'N_i$$

$$N_j=UN_i \tag{4D}$$

$$G_i=H_0F_i'+H_1F_i$$

$$G_j=H_2F_i'+H_3F_i \tag{4E}$$

The paired equations 4(C) may be rewritten as the single matrix relation 3.

Revised Minimization Concept

Overemphasis of minimization techniques appears to be a tradition rather than an actual economical goal in the design of high-speed logic. For data processing in the megacycle range, proper matching, minimum delay, and fast rise time become more important in order to reduce the number of pulse-shaping amplifiers, followers, and inverters as far as possible.

Earlier analysis of diode logic brought to light the implications of passive multilevel logic.[6] With the availability of new manufacturing techniques for matrices for 2-level logic advanced methods for efficient design are required to achieve the optimum ratio of logical elements to amplifiers with a minimum of pulse distortion.

Two steps are necessary to minimize a logical network to represent given function matrices: 1. determination of preferred forms, and 2. elimination of redundancy.

PREFERRED FORM

It has been shown that any function may be written either as a sum of minterms N_i or as a product of maxterms Mj.[1] Both forms are directly related as follows:

$$\sum N_i=\Pi Mj \quad N_i'=Mj \quad i+j=2^p-1 \tag{5}$$

The matrix concept deals only with the coefficients of minterms which are ONE's. If such a matrix comprises more ZERO's than ONE's, then the function itself shall be subject to minimization. Otherwise the complement contains less terms and should be processed. As an elementary example the function $A+BC'$ is considered:

$$A \qquad 01010101$$

$$BC' \qquad 00110000$$

$$A+BC' \quad (01110101)=f \tag{6A}$$

$$(A+BC')'(10001010)=f' \tag{6B}$$

Three ZERO digits are contained in the f matrix of equation 6(A). Accordingly, the complement f' involves three terms only and should be processed.

In general, processing of a function f of p variables instead of the complement f', is preferable if the matrix comprises fewer ONE's than ZERO's or fewer than 2^{p-1} minterms. Eventually the basic operation of the available hardware will be the determining factor. Diode matrices and several other components necessarily lead to AND/OR logic as represented by minterm matrices f. Direct or resistor-coupled transistor logic is related to maxterms or the reversed minterm matrices.[1]

ELIMINATION OF REDUNDANCY

Minimization of a logical function requires the elimination of redundancy. Redundancy exists if two terms have dummy indexes which differ by only one bit of their binary expression, that is, if

$$i-i^*= /2^k/ \qquad k=0, 1, 2, 3\ldots \tag{7}$$

It is believed that this sufficient condition has not been stated previously in such explicit form and it leads to several routines for minimization.

If different pairs of minterms, $i_1i_1^*i_2i_2^*$, satisfy the simultaneous equations 8(A) and (B) for different values of k, then the redundancy is of a higher order and the variables related to the values of k are redundant.

$$i_1-i_1^*=i_2-i_2^*=2^{k_1} \tag{8A}$$

$$i_1-i_2=i_1^*-i_2^*=2^{k_2} \tag{8B}$$

From a hardware point of view paired minterms required a single gate instead of two gates if a single variable is redundant. If both equations 8(A) and 8(B) are satisfied, the four gates required may be substituted by a single gate with all but the two redundant inputs.

REDUNDANCY IN FUNCTION MATRICES

It is possible to make a systematic search for redundant terms in the following manner.

If the dummy index corresponding to ONE in the function matrix contains at least one bit which is ZERO (when expressed in binary form), then permuting these ZERO's to ONE's one at a time, will indicate the paired indexes i^* immediately. It is necessary only to check that for any of these paired indexes, the coefficients in the function matrix are ONE.

It has been emphasized previously that each bit of the dummy index is related to a certain variable by a weight of a power of 2. Therefore a suitable matrix can be established to indicate the redundant variable for any pair of minterms. For an elementary case of three variables, A, B and C, the matrix $[r_3]$ may be derived as follows. If there is a coefficient $F_0=1$ corresponding to $i=000(b)$, then the pairs to be investigated are $i^*=001$; 010; $100(b)=1$; 2; 4. Weighting $A=1$, $B=2$, $C=4$, allows each variable to be placed in the appropriate column to indicate directly the redundant variable.

$$i=\begin{array}{cccccccc}0&1&2&3&4&5&6&7\end{array}$$

$$i(b)=\left\{\begin{array}{cccccccc}0&0&0&0&1&1&1&1\\0&0&1&1&0&0&1&1\\0&1&0&1&0&1&0&1\end{array}\right\}\begin{array}{l}\ldots C\\\ldots B\\\ldots A\end{array}$$

$$[r_3]=\begin{bmatrix}1&A&B&&C&&&\\&1&&B&&C&&\\&&1&A&&&C&\\&&&1&&&&C\\&&&&1&A&B&\\&&&&&1&&B\\&&&&&&1&A\\&&&&&&&1\end{bmatrix}\tag{9}$$

In order to determine if any redundancy exists, only a matrix multiplication of $[r]\times[f]=[rf]$ is required. This yields the redundant variables. Examples are contained in the Appendix II together with a matrix $[r_6]$ for minimization of six (or more) variables.

General Rules for Minimization of Function Matrices

EXISTENCE OF REDUNDANCY

By definition, redundancy exists in the minterms corresponding to the terms of a row of the $[rf]$ matrix whenever a variable occurs in this row.

ROW REDUNDANCY

The redundancy expressed in a row is covered by other rows if every term in this row occurs in a column which includes at least one other term. Owing to the method used in deriving the rf matrix, a variable will always occur in the same column as a ONE in another row. In checking for redundancy of rows, it is therefore necessary only to check if the ONE in that row has a variable from another row in its column.

GENERAL PRINCIPLES

1. By definition, if a row includes a ONE and a variable, that variable is redundant in the pair of minterms corresponding to the appropriate columns.

$$\begin{array}{llll}
[f] & 00\boxed{\begin{array}{c}11\\1A\end{array}}0001 & 00000001 & \\
[rf] & & & BC'D'\\
 & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & & \text{(redundant row)}\\
 & \qquad 1 & \qquad D & ABC\\
 & & & \ldots\text{(redundant row)}\\
\cline{4-4}
 & & & B(C'D'+AC)
\end{array}$$

(A is redundant in $A'BC'D'+ABC'D'=BC'D'$)

2. More than one variable in a given row indicates a NOT condition of these variables multiplied by the minterm of any of the occupied columns with these variables deleted. This is self-evident as shown in the following example:

$$\begin{array}{lll}
[f] & 01010100 & \\
[rf] & \boxed{1\ B\ C} & A(B'+C')\\
 & \quad 1 & \\
 & \qquad 1 &
\end{array}$$

This indicates that B is redundant in $AB'C+ABC'$, and C is redundant in $AB'C'+AB'C$. This leaves $AC'+AB'=A(B'+C')$.

3. (n) variables in a given column indicate an OR condition of the permutations of these variables taken $(n-1)$ at a time, multiplied by the corresponding minterm with these variables deleted. This is also self-evident as follows:

$$\begin{array}{lll}
[f] & 00000001\ 00010101 & \\
[rf] & \begin{array}{|cccc|}\hline 1&&&D\\&1&&C\\&&1&B\\\hline&&&1\end{array} & \begin{array}{l}ABC\\ABD\\ACD\\\hline A(BC+BD+CD)\end{array}
\end{array}$$

from paragraph 1 of this section, this is equal to $ABC+ABD+ACD=A(BC+BD+CD)$

4. If a number of variables occur simultaneously in a column and a row, and if the significant variables of the minterms in each case are identical, then the redundant variables are redundant simultaneously for all the minterms in the appropriate columns.

$$\begin{array}{lll}
[f] & 00000101\ 00000101 & \\
[rf] & \boxed{1\ B\qquad D} & AC(B'+D')\\
 & \quad\boxed{\begin{array}{ccc}1&&D\\&1&B\\&&1\end{array}} & AC(B+D)\\
\cline{3-3}
 & & AC
\end{array}$$

$AC(B'+D')$ (paragraph 2), $AC(B+D)$ (paragraph 3) $AC(B'+D')+AC(B+D)=AC$.

RULES AND PROCEDURE

In order to achieve optimum minimization, the following must be carried out:

1. Search for columns and rows combining the same variables and apply rule I.

Rule I. If the same variables occur simultaneously in a column and in a row, and if at least one of the columns containing the terms of the row under consideration, and one of the rows containing the terms of the column under consideration, include a common ONE, then the minterms corresponding to any relevent column omitting all the variables involved should be written down as part of the minimized expression.

2. The next step is the elimination of redundant rows by the application of rule II.

Rule II. Starting from the topmost row, rows which have not been involved in rule I and in which the ONE occurs in a column which includes a variable from another row, should be deleted, one at a time, and the result of these deletions taken into account in deleting further rows.

3. The next steps are to apply rules III, IV, and V in succession, and finally to add together all the expressions obtained.

Rule III. If n variables occur in a column, the reduced term for the minimized expression consists of an OR condition of the permutations of these variables taken $(n-1)$ at a time, multiplied by the corresponding minterm with these variables omitted.

Rule IV. If two or more variables occur in a row, the reduced term for the minimized expression consists of a NOR condition for these variables multiplied by the minterm of any of the occupied columns with these variables omitted.

Rule V. If only one variable occurs in a row, then the minterm corresponding to the column containing the variable should be written down in the minimized expression, omitting the variable in question.

Conclusions

A matrix equation, called expansion theorem, has been derived to compute the function matrix or truth table of a logical network expanded by adding a new input representing another independent variable. Its particular application is to determine the function matrix of combined moduli similar to the numerous transistor or magnetic blocks used in modern digital systems.

The inversion of this expansion theorem leads to a matrix routine for minimizing logical networks whereby the product of function matrix and a standard redundancy matrix gives directly the terms containing redundant variables.

It is believed that the methods offer considerable advantages in the design of complex logical networks as experienced in code translators, digital displays, or printing devices.

Application of the methods are shown in the appendixes for problems of various complexity.

Appendix I. Developing a Function Matrix by the Expansion Theorem

As an example of the use of the expansion theorem the function matrix of a chain of NOR circuits is calculated. Each NOR circuit has two inputs, the first circuit having the inputs *A* and *B*. Its output is connected to the following NOR circuit and a variable *C* is introduced as the other input.

Referring to Table I, the function matrix for the first level may be written immediately from Y_1:

$[f_{(2)}]=(1000)$

To determine the output at the second level of a NOR circuit requires application of (3), where $[g]=(1000)$

$$\begin{bmatrix}G_t\\ G_{t+4}\end{bmatrix}=\begin{bmatrix}1&0\\0&0\end{bmatrix}\times\begin{bmatrix}0111\\1000\end{bmatrix}=\begin{bmatrix}0111\\0000\end{bmatrix}$$

$[f_{(3)}]=(01110000)$

Following the third level with another variable as input together with $f_{(3)}$, the function matrix $f_{(4)}$ may be obtained similarly:

$$\begin{bmatrix}G_t\\ G_{t+8}\end{bmatrix}=\begin{bmatrix}1&0\\0&0\end{bmatrix}\times\begin{bmatrix}10001111\\01110000\end{bmatrix}=\begin{bmatrix}10001111\\00000000\end{bmatrix}$$

$[f_{(4)}]=(10001111\ 00000000)$

It will be noted that the next higher function matrix $[f_{p+1}]$ is

$[f_{(p+1)}]=[f'_{(p)},0_{(p)}]$

Accordingly:

$f_{(5)}=(01110000\ 11111111\ 00000000\ 00000000)$

$f_{(6)}=(10001111\ 0000000\ 11111111\ 11111111\ 0000000\ 0000000\ \ldots 00)$

Table II. Redundancy Matrix [r_6]

	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7
	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
00	1	a	b		c				d								e								.								f								.								.								.							
01		1		b		c			.	d							.	e							.								.	f							.								.								.							
02			1	a			c		.		d						.		e						.								.		f						.								.								.							
03				1				c	.			d					.			e					.								.			f					.								.								.							
04					1	a	b		.				d				.				e				.								.				f				.								.								.							
05						1		b	.					d			.					e			.								.					f			.								.								.							
06							1	a	.						d		.						e		.								.						f		.								.								.							
07	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	d	.	.	.	.	.	.	.	e	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	f	.								.								.							
10									1	a	b		c				.								e								.								f								.								.							
11										1		b		c			.								.	e							.								.	f							.								.							
12											1	a			c		.								.		e						.								.		f						.								.							
13												1				c	.								.			e					.								.			f					.								.							
14													1	a	b		.								.				e				.								.				f				.								.							
15														1		b	.								.					e			.								.					f			.								.							
16															1	a	.								.						e		.								.						f		.								.							
17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	e	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	f	.								.							
20																	1	a	b		c				d								.								.								f								.							
21																		1		b		c			.	d							.								.								.	f							.							
22																			1	a			c		.		d						.								.								.		f						.							
23																				1				c	.			d					.								.								.			f					.							
24																					1	a	b		.				d				.								.								.				f				.							
25																						1		b	.					d			.								.								.					f			.							
26																							1	a	.						d		.								.								.						f		.							
27	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	d	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	f	.							
30																									1	a	b		c				.								.								.								f							
31																										1		b		c			.								.								.								.	f						
32																											1	a			c		.								.								.								.		f					
33																												1				c	.								.								.								.			f				
34																													1	a	b		.								.								.								.				f			
35																														1		b	.								.								.								.					f		
36																															1	a	.								.								.								.						f	
37	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	f
40																																	1	a	b		c				d								e								.							
41																																		1		b		c			.	d							.	e							.							
42																																			1	a			c		.		d						.		e						.							
43																																				1				c	.			d					.			e					.							
44																																					1	a	b		.				d				.				e				.							
45																																						1		b	.					d			.					e			.							
46																																							1	a	.						d		.						e		.							
47	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	d	.	.	.	.	.	.	.	e	.							
50																																									1	a	b		c				.								e							
51																																										1		b		c			.								.	e						
52																																											1	a			c		.								.		e					
53																																												1				c	.								.			e				
54																																													1	a	b		.								.				e			
55																																														1		b	.								.					e		
56																																															1	a	.								.						e	
57	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	e
60																																																	1	a	b		c				d							
61																																																		1		b		c			.	d						
62																																																			1	a			c		.		d					
63																																																				1				c	.			d				
64																																																					1	a	b		.				d			
65																																																						1		b	.					d		
66																																																							1	a	.						d	
67	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	d
70																																																									1	a	b		c			
71																																																										1		b		c		
72																																																											1	a			c	
73																																																												1				c
74																																																													1	a	b	
75																																																														1		b
76																																																															1	a
77	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1

Appendix II. Minimization Chart and [r]-Matrix for Six or More Variables

Table II contains the [r]-matrix for six variables and may be expanded as follows:

row 00, column 100	variable *G*	r_7
row 00, column 200	variable *H*	r_8
row 00, column 400	variable *J*	r_9
row 00, column 1000	variable *K*	r_{10}
row 00, column 2000	variable *L*	r_{11}

Similar to the matrix [r_6] the variables continue diagonally with appropriate submatrices [r_6].

In order efficiently to handle problems of higher complexity the matrices are written for those columns only in which the function matrix has its coefficients $F_t=1$. At the top of the matrix the dummy indexes are listed in octal and/or decimal notation similar to appendix III, problem 3. Then from the matrix [r_6] or its expansion for high orders the appropriate binary products of R_tF_t are formed and placed properly.

For reference purposes the rules applicable for each case have been listed in the problems in Appendix III.

Appendix III. The Solution of Minimization in Logical Design

In order to facilitate a direct comparison of this method with other contemporary attempts towards the same goal, the following examples which have been taken from recent publications by other authors, are given.

Problem 1.[7] Translating the valid combinations into a function matrix yields

[f]	(00011101 10011001)	Significant terms	Rule
[rf]	1 D	AB	I
	1A D	$B'CD'$	IV, II
	1 B		II
	1 D		I
	1 C	$A'B'D$	V
	1 C		I
	1		II
	1		II

The minimized form is the sum of the significant terms and may be further reduced by conventional methods $AB+B'CD'+A'B'D=AB+B'(CD'+A'D)$.

Problem 2.[8] To determine the significant form of a 4-variable case with the truth table $i=0, 4, 6, 7, 10, 14, 15, 17$ in octal notation for $F_1=1$:

[f]	(10001011 10001101)	Significant terms	Rule
[rf]	1 C D	$A'B'$	I
	1 D		I
	1A	BCD'	V
	1 D		II
	1 C		I
	1A		II
	1 B	ACD	V
	1		II

Problem 3.[9] The dummy indexes i of the minterms are given in decimal notation as $i=0, 1, 2, 3, 7, 14, 15, 22, 23, 29, 31$. In obtaining the function matrix the columns are written in decimal (d) and octal (o) notation and limited to the columns containing $F_i=1$.

$$
\begin{array}{lll}
i(d) & 00000 & 112223 \\
 & 01237 & 452391
\end{array}
$$

$$
\begin{array}{lll}
i(o) & 00000 & 112233 \\
 & 01237 & 676757
\end{array}
$$

[f]	(11111 111111)	Significant terms	Rule
[rf]	1AB	$C'D'E'$	I
	1 B		I
	1A		I
	1C		II
	1 D E	ABC	I
	1A	$BCDE'$	V
	1 E		I
	1A	$BCD'E$	V
	1 D		I
	1B	$ACDE$	V
	1		I

The column of significant terms is equivalent to the minimum sum.

References

1. Matrix Analysis of Logical Networks, E. J. Schubert. *AIEE Transactions*, vol. 77, pt. I, Mar. 1958, pp. 10–13.

2. An Algebraic Theory for Use in Digital Computer Design, E. C. Nelson. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., vol. EC-3, no. 3, Sept. 1954, p. 12.

3. Mechanized Reasoning, D. M. McCallum, J. B. Smith. *Electronic Engineering*, London, England, Apr. 1951, p. 126.

4. Transistor NOR Circuit Design, William D. Rowe, George H. Royer. *AIEE Transactions*, vol. 76, pt. I, July 1957, pp. 263–67.

5. Magnetic Switching Circuits, R. C. Minnick. *Journal of Applied Physics*, New York, N. Y., vol. 25, no. 4, Apr. 1954.

6. Engineering Multistage Diode Logic Circuits, B. J. Yokelson, W. Ulrich. *AIEE Transactions*, vol. 74, pt. I, Sept. 1955, pp. 466–75.

7. Minimization of the Partially Developed Transfer Tree, M. P. Marcus. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, vol. EC-6, no. 2, June 1957, p. 94.

8. An Algorithm for Determining Minimal Representations of a Logic Function, B. Harris. *Ibid.*, p. 105.

9. Minimization of Boolean Functions, E. J. McCluskey, Jr. *Bell System Technical Journal*, New York, N. Y., 1956, p. 1427.

Volt-Second Transfer Efficiency in Fast-Response Magnetic Amplifiers

Part II. N^2/R as a Design Parameter

T. J. PULA
ASSOCIATE MEMBER AIEE

G. E. LYNN
MEMBER AIEE

J. F. RINGELMAN
ASSOCIATE MEMBER AIEE

THIS PAPER presents the final portion of a study to extend the volt-second control approach in considering the effects of signal circuit resistance and nonideal flux-current loops for fast-response magnetic amplifiers.[1,2]

In the first portion of the investigation, volt-second transfer efficiency measurements were used to demonstrate the nature of N^2/R as a basic reactor parameter. This knowledge and the assumption of constant average incremental ampere-turns for complete reset were then used in deriving expressions which indicate the complementary relationship existing between volt-seconds and ampere-turns in magnetic amplifier control.[3]

The object of this concluding portion of the study is to substantiate the ideas already developed by correlating the measurements of incremental volt-second transfer efficiency from part I[3] with gain measurements for a specific fast-response magnetic amplifier.

Illustrative Fast-Response Amplifier

The fast-response amplifier chosen to demonstrate this correlation has been under development by the authors for several years and has been designated the Hybrid III amplifier. The name is in a sense generic and is derived from a somewhat hopeful assumption that the amplifier weds a first stage, which is (to a first-approximation) ampere-turn controlled, to a second stage, which can be considered largely volt-second controlled. That neither of these conditions is ever totally satisfied in the amplifier is demonstrated in the paper.

CIRCUIT OPERATION

The circuit for discussion, shown, in Fig. 1, is a 2-stage push-pull amplifier with d-c reversible polarity output. The second stage is a push pull arrangement of Ramey-type circuits. Angle volt-second control of each half of this stage is provided by a Logan-type circuit.[4] The basic design requirement for this type of control is that the "driving ampere-turn" value for the gate circuit of the first stage must be less than the "threshold ampere-turn" value for the control circuit of the second stage.[4] When this requirement is satisfied, substantially all of the volt-seconds from E_{ac} are absorbed by the first stage reactors until they fire. The amount of E_{ac} volt-seconds thus absorbed, furthermore, is determined by the initial flux condition of the first stage cores as set by the control provided to the Logan circuit. After these first stage cores fire, the remaining volt seconds from E_{ac} are applied to the second-stage control winding in series with all interstage voltage drops.[5,6] It is the inevitable existence of these voltage drops which precludes the possibility of true volt-second control.

Paper 58-173, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted October 30, 1957; made available for printing August 7, 1958.

T. J. Pula, G. E. Lynn, and J. F. Ringelman are with the Westinghouse Electric Corporation, Baltimore, Md.

Vol.
Pt. 1
1959

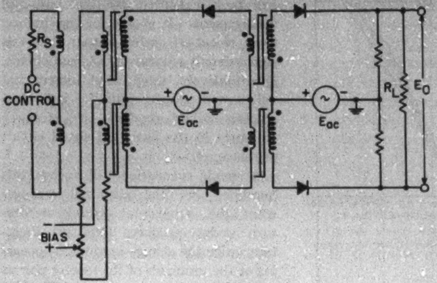


Fig. 1 (left). Circuit of Hybrid III fast-response magnetic amplifier

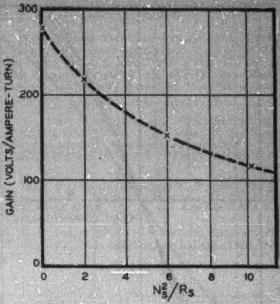


Fig. 2 (right). Volts per ampere-turn as a function of N²/R for a Hybrid III

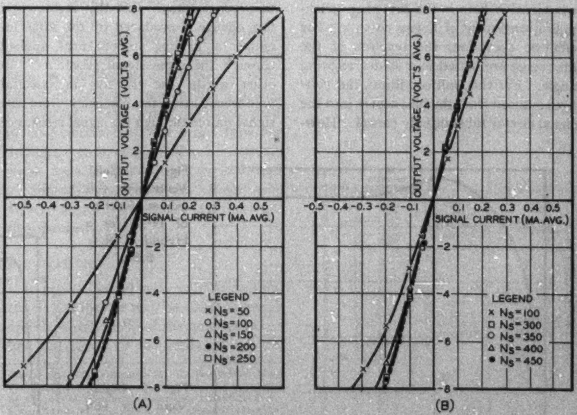
For fast-response design, E_{ac} must exceed nE_{ac}' by a considerable percentage in order to prevent interstage current while the first stage cores are resetting. Therefore, the interstage voltage drops may not be made negligible here and thus even during this time interval determinate volt-second control is not desirable even if it were obtainable. The coupling can be designed in such a way, however, that the transfer efficiency in this interstage coupling loop remains fairly constant while the first-stage cores are allowed to reset without producing appreciable interstage current. Under these design conditions, this second stage merely provides an additional volts/volt gain and the over-all volts/NI gain is essentially a function of the first stage core configuration.

During the prefiring interval of the gating half-cycle for both first-stage cores, the voltages induced into the control circuit are in opposition and approximately equal. However, for any condition other than null, one core fires before the other. Therefore, during the interval in the gating half-cycle after one core has fired, the continuation of flux change in the other core induces volt-seconds into the control circuit. Because of the push-pull, 90-degree bias, reversible d-c character of this circuit, the induced voltage during this period always causes the instantaneous i_s to exceed E_s/R_s. During the period when both cores are being reset, the magnitude of the control current is less than E_s/R_s because of the reactor resetting voltage. However, the reset $\Delta\phi$ obtained (and therefore the ΔE_0) is a function of the magnitude of control current during reset only and not of the average control current. And if N^2/R is increased, the percentage difference between the resetting current and the average control current is increased. Therefore, the V/NI_{avg} gain of such an amplifier decreases as N^2/R increases.

Application of N²/R to Design

The volts/NI gain versus N^2/R for an amplifier of this type is shown in Fig. 2. For convenience, this curve was obtained by varying R_s since the fundamental nature of N^2/R has already been established in part I.[3] The N_c^2/R_c value for which the volts per ampere-turn gain is equal to one half the gain obtained when N_c^2/R_c is equal to zero (extrapolated from the curve of Fig. 2) was picked off as equal to 7.8 (adding 0.50 for bias). The value of gain at $N_c^2/R_c=0$ is designated as $(G_v/NI)_0$. Two practical values were selected to represent minimum allowable R_s (i.e., 8.6k, 3k), and the optimum N_s calculated for each. Volts/NI gain curves were then run for each R_s value by varying N_s. These curves are shown in Figs. 3 and 4, and the resulting calculated curves of volts/volt gain versus N^2/R in Figs. 5 and 6. It can be seen that a maximum volts/volt gain does exist and that it occurs at a value of N_c^2/R_c corresponding to the value yielding G_v/NI equal to 1/2 $(G_v/NI)_0$. This confirms the relationship between volt-seconds and ampere-turns in magnetic amplifier control predicted in part I.

The correlation between the N^2/R producing maximum volts/volt gain and the N^2/R producing 50% incremental volt-second transfer efficiency in the individual reactors employed in the first stage of the amplifier will now be discussed. In part I of this study, it was predicted from mathematical considerations that the maximum volts/volt gain for an amplifier whose input configuration reduces to a Logan circuit should be obtained for an N^2/R corresponding to 50% incremental volt-second transfer efficiency. This will



(A)

(B)

Fig. 3. Output voltage as a function of signal current for a Hybrid III ($R_s = 8.6K$)

A—Low N_s
B—High N_s

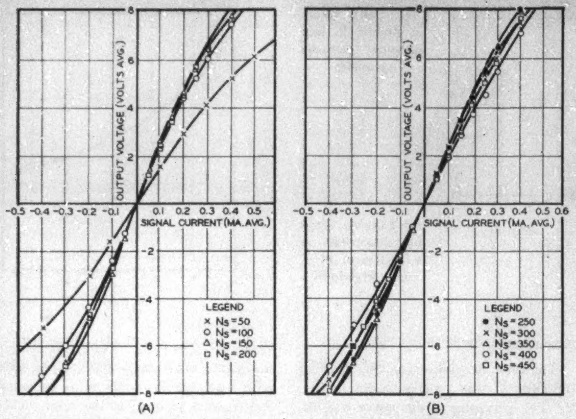


Fig. 4. Output voltage as a function of signal current for a Hybrid III ($R_s=3.1K$)

A—Low N_s
B—High N_s

remain true for a 2-stage amplifier if the second stage contributes only a volts/volt gain. In order to check the correlation under investigation, a closer examination of the circuitry in the Hybrid III magnetic amplifier of Fig. 1 is required.

The function of the second stage in merely producing additional volts/volt gain for this amplifier when interstage coupling current during reset of the first-stage cores has been suppressed by appropriate design, has already been discussed. The circuitry requiring closer examination here is the first-stage push-pull arrangement of Logan circuits. For efficient operation, the reactors of the first stage are biased to a 90-degree firing angle. For this null condition, the voltages induced from the bias source into the signal circuit substantially cancel. However, for a general signal condition, one of the input stage cores saturates before the other. For such a general operating condition, the bias source is still attempting to produce half of maximum flux change in each core on each reset half-cycle. Signal, however, is aiding bias in resetting one core but opposing bias on the other core. As contrasted with the null condition then, one of the cores resets by an amount $(\Delta\phi_b-\Delta\phi_x)$ and the other resets by the greater amount $(\Delta\phi_b+\Delta\phi_x)$. The effect of signal, therefore, is to produce $2\Delta\phi_x$. But this incremental flux change is produced by the same incremental driving ampere-turns applied through the same signal resistance. In other words, the effective incremental volt-second transfer efficiency from the signal source through the signal resistance has been enhanced by the push-pull arrangement. It is only for the limiting cases of cut-off or full output that this signal circuit reduces to a Logan circuit. Quantitatively, then, what incremental volt-second transfer ratio for individual reactors should produce 50% transfer efficiency in the push-pull signal circuit of this amplifier?

A simple calculation will answer this question. For the push-pull arrangement, the incremental driving ampere-turn product produces $2\Delta\phi_x$ and, therefore, twice the change in voltage appearing at the terminals of the reactor pair as would be produced across an individual reactor by this same signal current. Thus, if the incremental volt-second transfer efficiency obtained with a matched pair of reactors is 50% in the push-pull configuration, the incremental volt-second transfer efficiency measured in our special test circuit at the same N^2/R for one of these reactors should be 33.3%. (This follows immediately since the resistor reset voltage remains fixed while half of the voltage which appeared at the reactor-pair terminals disappears.)

With this clarification in mind, the correlation obtained can now be assessed. From measurements on the first stage cores, it can be seen from Fig. 5 of reference 3 that a signal circuit N^2/R of 7.5 (with a bias circuit $N^2/R=0.3$) yields an incremental volt-second transfer efficiency of approximately 32%. But this is the same control N^2/R (i.e., 7.8) which produces maximum volts/volt gain when reactors from two such cores are used in push-pull. This close agreement confirms the validity of the results derived in part I from theoretical considerations.

Conclusions

The incremental volt-second transfer efficiency characteristics of reactors composed of square-loop magnetic materials presented in part I have been applied to the design of a modern fast-response

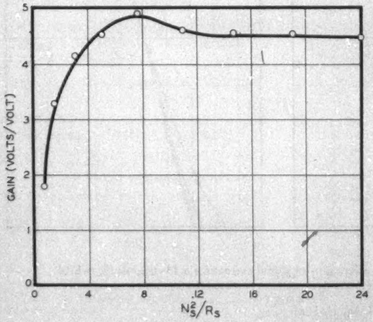
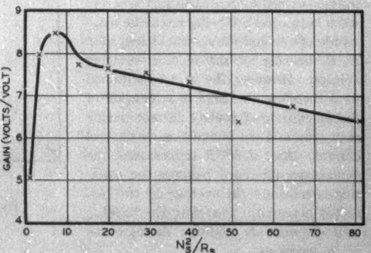


Fig. 5 (left). Volts per volt gain as a function of N^2/R for a Hybrid III (R_s=8.6K)

Fig. 6 (right). Volts per volt gain as a function of N^2/R for a Hybrid III (R_s=3.1K)



magnetic amplifier. Close correlation was obtained between the signal circuit N^2/R producing an effective incremental volt-second transfer efficiency of 50% and the N^2/R value yielding maximum volts per volt gain in the amplifier. This confirms the complementary relationship between volt-seconds and ampere-turns in fast response magnetic amplifier control developed in the previous paper from theoretical considerations. Experimental results show that the incremental volt-second transfer efficiency of a fast-response amplifier defines the amplification obtainable with a given configuration. This transfer efficiency is dependent upon the N^2/R of the control circuits, proving this parameter to be as basic to the amplification factor of a fast-response amplifier as it is to the gain-bandwidth product of the older, slower circuits such as the doubler or bridge type.

References

1. On the Mechanics of Magnetic Amplifier Operation, R. A. Ramey. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1214–22.

2. On the Control of Magnetic Amplifiers, R. A. Ramey. *Ibid.*, pp. 2124–28.

3. Volt-Second Transfer Efficiency in Fast-Response Magnetic Amplifiers, Part I. N^2/R and Control, T. J. Pula. *Ibid.*, vol. 77, pt. I, 1958 (Jan. 1959 section), pp. 861–67; also "Conference on Magnetic Amplifiers," *AIEE Special Publication T-98*, Aug. 1957, pp. 107–20.

4. A New Approach to Self-Saturating Magnetic Amplifiers, T. J. Pula, F. G. Timmel. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. XII, Apr. 1957, pp. 425–43.

5. Fast Response Magnetic Amplifiers, G. E. Hughes, H. A. Miller. *AIEE Transactions*, vol. 73, pt. I, Mar. 1954, pp. 69–75.

6. The Pulse Stretch Coupling Circuit, H. W. Patton. *Ibid.*, vol. 75, pt. I, Sept. 1956, pp. 377–79.

A Miniaturized Negative-Impedance Voice Repeater Employing Transistors

A. S. HOWELL
MEMBER AIEE

UNPRECEDENTED service demands on the telephone operating companies have emphasized the need for an economical means to extend transmission ranges and to permit the reduction of cable gages in local office areas. As a result, considerable interest has developed in the negative-impedance-type repeater, recently introduced and now finding wide application by the Bell System.

Current demands for such new facilities, however, often create serious space problems. This is particularly true when the required equipment must be installed in existing bays. As one solution to this problem, two miniaturized negative-impedance-type repeaters have been developed. These repeaters employ transistors as their active elements. They are entirely powered from a 48-volt central office battery, thus eliminating any need for bulky power supplies. Modern plug-in printed-wiring-board techniques have been successfully applied in the physical make-up of these repeaters. This makes for compact packaging.

General Description

The type *541* repeaters to be discussed are of the series and the series-shunt type. Their electrical function is to provide transmission gain when inserted in a 2-wire telephone line. Provisions for gain adjustment are adaptable to all line facilities normally found in the telephone outside plant. The circuit design is such as to minimize changes in gain over the range of expected environmental conditions.

The mechanical design includes separate, but companion, plug-in series-type and shunt-type repeater packages. Each unit is of the same over-all size, dimensionally restricted to shelf-mount up to eight repeaters on a standard 19-inch relay rack, within a 3½-inch vertical mounting space. Thus 208 repeater units may be placed in one 9-foot bay of a standard rack.

The series repeater consists of three functional parts: an open-circuit stable negative-impedance converter, a gain-adjusting network, and a line-coupling transformer. The shunt repeater, used in conjunction with the series type for additional gain, consists of a short-circuit stable negative-impedance converter and a gain-adjusting network. Both converters employ transistors as their active elements.

Application

In general, based upon stability factors, the series repeater may be used to recover approximately 50% of the nonrepeatered line loss, up to 10 db (decibels), without singing. However, the series repeater effectively operates on the series constants of the line and introduces an impedance irregularity at the point of insertion. This produces reflections and results in echo effects. The corresponding circuit transmission derating is proportional to the repeater gain and the length of the circuit. While such derating may be acceptable on local circuits it can become serious on long toll-connecting trunks and the use of the series-shunt repeater combination will be required. Such a combination may be used to recover up to approximately 85% of the nonrepeatered line loss and the reflection effects are appreciably reduced.

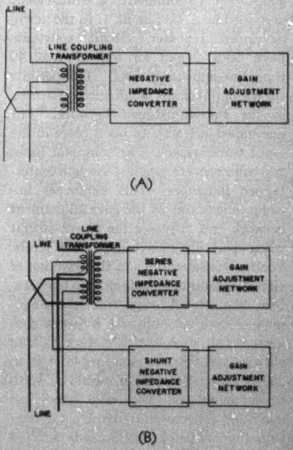

LINE
LINE COUPLING TRANSFORMER
NEGATIVE IMPEDANCE CONVERTER
GAIN ADJUSTMENT NETWORK
(A)
LINE
LINE COUPLING TRANSFORMER
SERIES NEGATIVE IMPEDANCE CONVERTER
GAIN ADJUSTMENT NETWORK
SHUNT NEGATIVE IMPEDANCE CONVERTER
GAIN ADJUSTMENT NETWORK
LINE
(B)

Fig. 1. Basic elements

A—Series repeater
B—Series-shunt repeater

Paper 57-1097, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted August 30, 1957; made available for printing August 6, 1958.

A. S. Howell is with the Stromberg-Carlson Company, Rochester, N. Y.

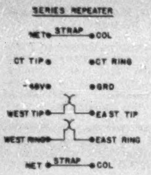

SERIES REPEATER
SHUNT REPEATER
NET
STRAP
COL
EM
CT TIP
CT RING
-48V
GRD
WEST TIP
EAST TIP
WEST RING
EAST RING

Fig. 2. Connector plan

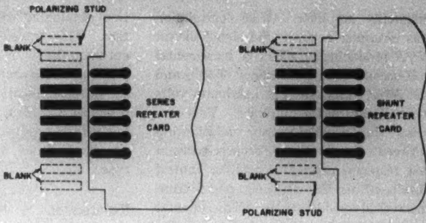
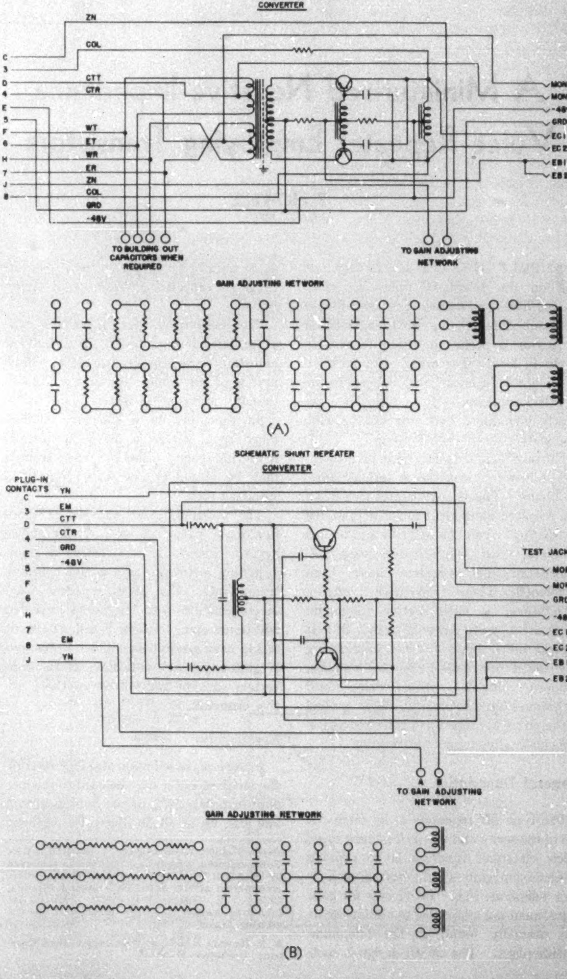

POLARIZING STUD
BLANK
SERIES REPEATER CARD
SHUNT REPEATER CARD

Fig. 3 (right). Polarizing plan

Fig. 4 (below). Schematics

A—Series repeater
B—Shunt repeater


SCHEMATIC SERIES REPEATER
CONVERTER
ZN
COL
CTT
CTR
WT
ET
WR
ER
GRD
-48V
MON
EC1
EC2
EB1
EB2
TO BUILDING OUT CAPACITORS WHEN REQUIRED
TO GAIN ADJUSTING NETWORK
GAIN ADJUSTING NETWORK
(A)
SCHEMATIC SHUNT REPEATER
CONVERTER
PLUG-IN CONTACTS
YN
EM
TEST JACKS
A
B
(B)

Use of Transistors

Junction transistors possess desirable properties for use in the negative-impedance circuit to be described. As has been pointed out,[1] their conversion properties in this application are stable and the conversion factor is substantially controlled by the alpha value. The present life expectancy of transistors compares favorably with that of vacuum tubes.

As the state of the art progresses, it is expected that the ideal of practically infinite life may be attained. Coupled with the obvious advantages of economy of space and required power consumption, the choice of transistors in this equipment appears well justified.

Discussion

Series-Repeater Circuit

The series-converter circuit is an adaptation of a common base arrangement described by J. G. Linvill.[1] In the newly developed repeater, circuit constants have been chosen for operation at 40 times line-impedance level. This permits the use of physically small mica capacitors in the gain-adjustment network. The line-impedance step-up takes place in the input transformer. The transistor bases and collectors are supplied through center-tapped inductors. This minimizes impedance shunting of the gain adjustment network and increases the power output when operating at high-impedance levels. Further considerations in the selection of circuit constants include proper temperature bias stabilization, and the maintenance of a substantially constant conversion factor both with respect to load changes and to changes in frequency from 250–3,500 cps (cycles per second). With these circuit constants established, the operating collector currents are less than 1 ma (milliampere) using p-n-p-type *2N44* or similar transistors.

Several considerations are involved in the design of the associated line-coupling (input) transformer. The series-aiding impedance of the line windings must be of

Vol.
Pt. 1
1959

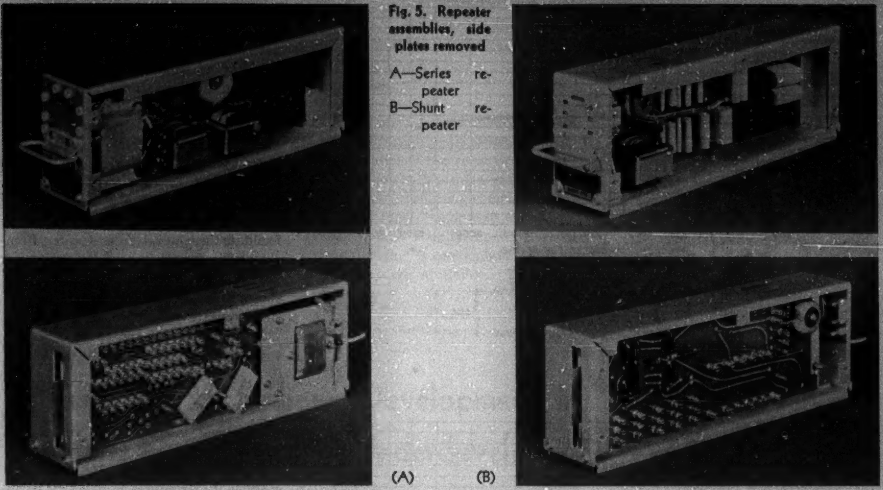


Fig. 5. Repeater assemblies, side plates removed

A—Series repeater
B—Shunt repeater

sufficient value to prevent excessive shunt loss at the lowest voice frequency to be amplified. This value must be maintained with up to 130 ma direct current in the line windings, based upon the possibility of two simultaneous telegraph signals at 65 ma, east and west on composited legs. The d-c resistance and the impedance to ringing frequencies must be sufficiently low for negligible restriction on the loop-signaling range. Line-winding balance must be capable of meeting standard longitudinal balance requirements.

Resistive and reactive components introduced by the input transformer must be compensated for ahead of the network if the gain adjustment is to be directly dependent upon the transformed line impedance. Loss factors contributing to these components are minimized through the use of a laminated core of high-permeability magnetic material in the transformer design. The inclusion of sectionalized windings reduces troublesome distributed capacitance. An interwinding electrostatic shield is provided to minimize stray capacitance couplings. The effective series resistance is substantially independent of frequency over the operating range and is compensated for by a resistor in series with the gain-adjusting network.

Shunt-Repeater Circuit

The balanced circuit of the shunt converter is basically similar to the series-repeater circuitry except that the positions of the input and network terminals are interchanged. No impedance-level transformation is involved. The negative conversion factor, somewhat less than minus alpha, is effectively brought close to -1 by addition of low-value resistances in series with the line-blocking capacitors. When shunt-input resistances are used to supply the collectors and their value is made sufficiently high to obtain negligible shunt-input loss, the supply voltage drop results in insufficient collector voltage to maintain the required operating conditions. For this reason the collectors are supplied through a center-tapped inductor. A high value of inductance is used in order to reduce the shunt losses at ringing frequencies. A shunt capacitor across this inductor limits the converter response at high frequencies.

Connection Facilites

The basic elements of the series repeater are illustrated in Fig. 1(A) and those of the series–shunt combination in Fig. 1(B). In the latter figure, center taps are provided on the two line windings of the coupling transformer and the shunt repeater elements are added in the bridged-T configuration.

A compact connector arrangement results from the application of a plug-in card including six paris of double-side printed contacts. The associated receptacle includes short-circuiting spring for the maintenance of line continuity after the card is removed. The gain-adjusting network is strap-connected to the converter circuitry at the receptacle. It is automatically disconnected when the repeater is removed for insertion into an associated test set. This plan is shown in Fig. 2.

The mounting shelf may be wired for any combination of series and shunt units by employing similar receptacles. Transposition of series and shunt repeat-

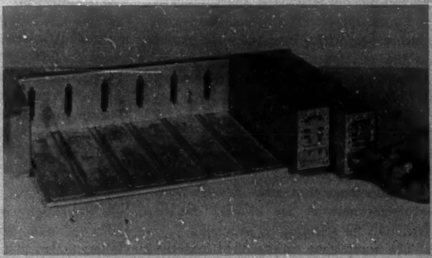


Fig. 6. Shelf-mounted repeater assembly

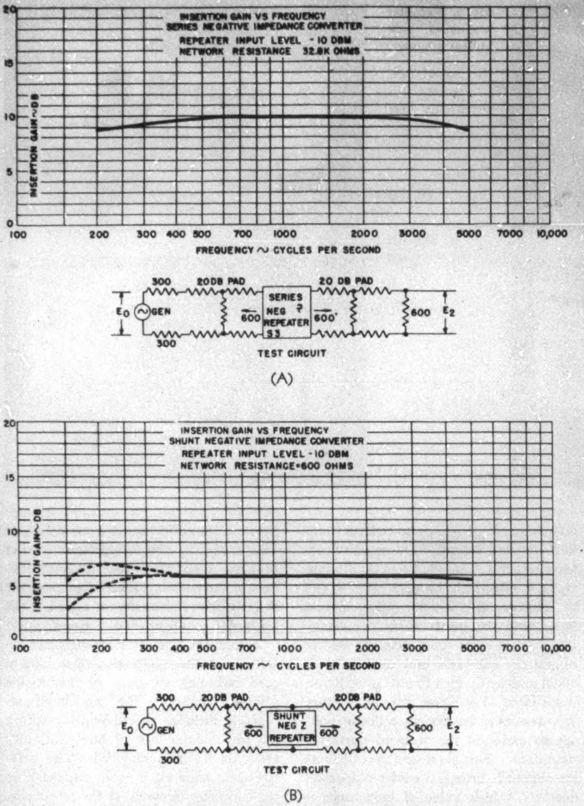


Fig. 7. Insertion gain versus frequency characteristic

A—Series repeater B—Shunt repeater

ers after the positions are established is prevented in the following manner: The receptacles include blank recesses at top and bottom equivalent to those occupied by two pairs of terminals; the series card is keyed opposite to the shunt-repeater card and polarizing insert clips are placed in the extreme upper and lower blank recesses as required. This arrangement is illustrated in Fig. 3.

The complete circuit for the series repeater is shown in Fig. 4(A), and that of the shunt repeater in Fig. 4(B). The graded values of resistance, capacitance and inductance that make up the gain-adjusting network are brought out to terminals for selection by strapping.

The component arrangement and the circuitry is readily adaptable to the application double-side printed-wiring cards. The network strapping terminals are mounted on one side of the card and substantially all components on the opposite side. The input transformer in the series repeater is mounted separately. These cards are fabricated from glass epoxy to minimize the possibility of cracking during the riveting of the more than 40 strapping terminals. They are copper-clad for printed wiring. The plug-in contacts are nickel-rhodium plated to minimize wear.

PACKAGING

The plug-in cards are enclosed in aluminum housing with removable side plates for accessibility of network-strapping terminals and components. This also results in effective voice-frequency shielding of adjacent repeaters. Base-mounted brackets support the printed boards at each end. The line-coupling transformer is separately mounted on one base-supported bracket. Insulated pin jacks for signal monitoring and for the measurement of collector voltages are located on the front face of the enclosures. Fold-flat pull handles are provided to facilitate the removal of the repeaters from the shelf. Two views of the series repeater, with side plates removed, are shown in Fig. 5(A). Similar views of the shunt repeater are shown in Fig. 5(B).

A front view of the mounting shelf with inserted repeaters is shown in Fig. 6.

PERFORMANCE

A typical insertion gain versus frequency characteristic for the series repeater is shown in Fig. 7(A). In this example the repeater is adjusted for 10-db gain at 1 kc when working between 600-ohm pads, each of 20-db attenuation.

A similar characteristic for the shunt repeater is illustrated in Fig. 7(B), in this case the repeater is adjusted for 6-db gain at 1 kc.

Summary of Characteristics

541 A21 SERIES REPEATER

Power supply = 44–54 volts, d-c
Power consumption = 0.6 watt at nominal 48-volt d-c
Power-output level = +10 dbm (db referred to one milliwatt), maximum
Total harmonic distortion, per cent:

Power Output	6-Db Gain	10-Db Gain
+0 dbm	1.4	1.9
+5 dbm	2.1	2.9
+10 dbm	2.6	4.0

Operating ambient temperature = 0 C (degrees centigrade) to +55 C
Gain change at maximum ambient temperature = +0.2 db maximum
Voice-frequency response = +1.0 db, −2.0 db, 250–3,500 cps (referred to 1-kc value)
Nominal maximum gain (resistive terminations) = 10 db
Noise level (*F1A* weighting) = less than 15 dba (db adjusted)
Over-all dimensions = $3^1/_{16}$ inches high x 2 inches wide x $9^1/_2$ inches long

541 A31 SHUNT REPEATER

Power supply = 44–54 volts, d-c
Power consumption = 1 watt at nominal 48-volt d-c

Power-output level = +10 dbm, maximum
Total harmonic distortion:

Power Output	Per Cent Distortion at 6-Db Gain
+0 dbm	0.032
5 dbm	0.25
+10 dbm	1.1

Operating ambient temperature = 0 C to +55 C
Gain change at maximum ambient terminations = +0.4 db, maximum
Voice-frequency response = ±1.0 db, 250–3,500 cps (referred to 1-kc value)
Nominal maximum gain (resistive terminations) = 6 db
Noise level (*F1A* weighting) = less than 15 dba
Over-all dimensions = $3^{1}/_{16}$ inches high x 2 inches wide x $9^{1}/_{2}$ inches long

Conclusions

The described designs illustrate a plan of miniaturization made possible by circuit application of the versatile transistor in conjunction with printed wiring. The plug-in units constitute compact, lightweight but rugged equipment enclosures. They are readily manufactured. Over-all circuit shielding is accomplished and all components are accessible. This same plan also may prove advantageous in the packaging of other equipments.

It is believed that this equipment will prove of value in meeting the requirements of telephone companies for an economical means to extend transmission range within their service areas.

References

1. Transistor Negative Impedance Converters, J. G. Linvill. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, no. 6, 1953, pp. 725–29.

2. Negative Impedance Telephone Repeaters, J. L. Merrill, Jr., A. F. Rose, J. O. Smethrust. *Bell System Technical Journal*, New York, N. Y., vol. 33, no. 5, 1954, pp. 1055–92.

A Review of Telephone Development in Western Europe and in Underdeveloped Areas

G. D. WALLENSTEIN
MEMBER AIEE

THIS PAPER represents a look at telephone development around the world. Emphasis is placed on some outstanding European contributions to the art of telephone transmission. An appraisal of telephone growth potential in the lesser developed countries is also included.

Telephone Communications Development in Europe

During the last 80 years there has been much interchange of ideas between European and American communications people. Two world wars and increasing mobility have contributed in making this one world from the standpoint of technological know-how.

Nevertheless, there are outstanding characteristics and trends in modern European telephone transmission systems which are not encountered in North America. The reasons for such differences may be classified as follows:

1. Several relatively small countries want to foster their own system development and supply.
2. In many countries telephone service remains a luxury because of low average income.
3. Ideas practical for high-volume application have no appeal where the number of salable systems is low, a result of the foregoing conditions 1 and 2.
4. Separate national developers tend to use raw materials and techniques available in their country and thus avoid import.
5. War-time destruction made much telephone plant useless, thus invalidating a prime incentive for developing communications systems to work on old-style cables and wire lines.
6. The International Telephone and Telegraph Consultative Committee (CCITT) establishes recommendations for international backbone circuits in Europe. While this is an absolute necessity in view of the divergent national interests, it tends to restrict such nonconforming developments for the shorter route traffic as have appeared on the American market.

What are the more interesting European developments generated in this way? This paper describes some of the more important developments and ideas of recent years. It is based on the writer's own impressions and is not intended to be a complete list.

Carrier Frequency Cables for Large Number of Telephone Channels

In the industrial area of Western Europe, many cities require heavy telephone trunk routes of relatively short length. In the late 1930's high-grade paper-insulated cables and associated carrier systems similar to the Bell system's type *K* were introduced

Postwar development took advantage of improved insulating materials and manufacturing methods which resulted in a high-frequency balanced type of cable not used so far in North America.[1] Up to 120-channel transmission is now practical on each balanced pair in a multipair cable exclusively designed for carrier use. Noise and crosstalk control is such that good toll standards are met without compandors at the terminals. Transmission attenuation is much lower than that of paper-insulated cables as shown in Fig. 1.

These cable systems are the backbone of Western Germany's toll network, and are used in other countries of Western Europe.[2] Their exploitation has evolved from 24 to 48, then 60, and now 120 channels by progressive improvements.

Cable construction is in multiples of star quads. Typical sizes are 7-quad and 12-quad. One type provides 8 quads plus one coaxial tube in the center. The conductors are 1.3-millimeter (16-gage) copper. Fig. 2 shows the conductor arrangement in the 7- and 8-quad sizes. Fig. 3 illustrates a short length of the 7-quad size showing construction details.

Repeater sections of 12 miles with a section loss of about 60 db (decibels) at 252 kc have been standard for the 60-channel system. The transmission fre-

---

Paper 57-786, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted April 19, 1957; made available for printing September 22, 1958.

G. D. Wallenstein is with the Lenkurt Electric Company, San Carlos, Calif.

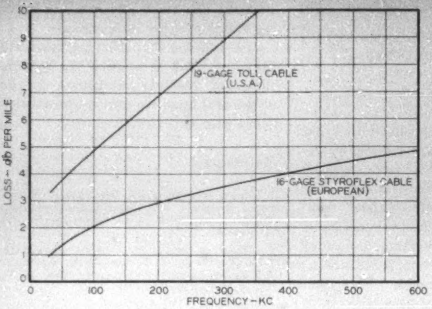


Fig. 1. Comparison of the attenuation characteristics of European manufactured Styroflex toll cable (16 gage) and American 19-gage toll cable. The European cable is usable up to about 600 kc. The American cable has a practical upper limit of less than 300 kc

quencies follow the CCITT recommended 12-channel group scheme. A 60-channel system contains five such groups in the following ranges:

12 . . . 60 kc
60 . . . 108 kc
108 . . . 156 kc
156 . . . 204 kc
204 . . . 252 kc

This scheme allows no space between groups, a fact which has made necessary the development of expensive group-transfer filters to suppress adjacent groups separated by only a few hundred cycles in frequency.

The introduction of a second 60-channel block over the same 12-mile repeater sections was made possible by the use of Styroflex-insulated cable types; see Fig. 3. (Styroflex is an oriented-molecule form of polystyrene.) A pair in this cable has about the same loss at 550 kc as paper-insulated pairs have at 250 kc. This made it possible to use a complete standard 60-channel supergroup in the range 312–552 kc on routes with the new cables. The gap of 60 kc between the two 60-channel allocations offers an unusually economical chance for group transfer.

Control of crosstalk is a major problem on these cables. The two directions of transmission are handled by two separate cables. Even so, within one cable sheath, the combinations of capacitive and electromagnetic couplings add up to

$$\frac{n}{2}(n-1)$$

where n is the number of pairs. For a 24-pair cable, this figure is 276. Despite very rigorously applied specifications for the manufacture of these cables, it is necessary to use crosstalk-balancing devices, one for each repeater section. The importance of capacitive balancing may be illustrated by the fact that as little as a 4-micromicrofarad capacitive coupling causes a far-end crosstalk-coupling loss no better than 70 db. With capacitor-balancing panels it is possible to increase the far-end crosstalk-coupling loss to about 75 db or better for all combinations over the length of one repeater section. These values apply at the highest transmitted frequency.

The question arises whether this technique is indeed more economical than the use of cheaper voice-frequency cables with compandored channels to offset crosstalk. The relative high cost of a precision-manufactured high-frequency cable and of balancing the pairs can be prorated over 120 channels, whereas no more than 24 channels can be operated on the compandored systems customary in the United States. On the other hand, in American systems the two directions can be handled by two pairs in the same cable sheath compared to separate pairs in two parallel cables for European systems. Hence for the European systems there is a higher first cost. However, it would appear that the five times greater channel capacity and the saving of two compandors in each channel gives the edge to the European-type cable system. This comparison applies to new cable installations only.

Even if costs were equal, the European preference would be in favor of non-compandored toll-quality transmission techniques. The CCITT has not so far accepted compandors as a permissible tool for meeting noise and crosstalk requirements. International circuits must be built up with compatible channel terminals at each end. If a compressor were used at one end, a corresponding tracking expandor would be needed at the other. This would be a further complication of international technical agreements, and would impose an economic penalty on that part of the circuit which uses transmission facilities not requiring the compandor advantage. The alternative is to connect circuits at voice frequencies so that both channel

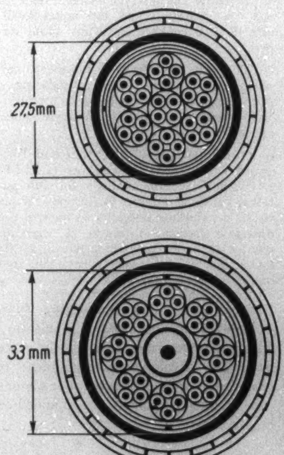


Fig. 2. Conductor and quad arrangements in 7- and 8-quad Styroflex cables. The 8-quad cable also contains a coaxial transmission line

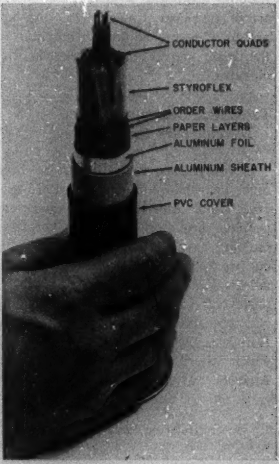


Fig. 3. Construction of latest type 7-quad cable. In addition to the quads, order wires are placed between outer layers of paper insulation

terminals would always be under control of one administration. However, this is a backward step and would again be felt as a hardship.

It is evident that these high-channel capacity balanced-pair cable-carrier systems are an excellent solution for Western Europe's telephone traffic problem. They fit into the already standardized group-transmission scheme, are economical to operate, and are capable of expansion with well-proved techniques.

SHORT-HAUL CARRIER SYSTEMS OF UNCONVENTIONAL DESIGN

The impact of CCITT recommendations has been a freezing of system transmission parameters such as 4-kc channel spacing, standardized 12-channel group frequencies, and connecting impedances. Officially these apply only to international circuits. Some manufacturers prefer to stay within these recommendations regardless of a system's field of application; others venture into nonconforming fields for short-haul transmission.

A large amount of research and economical analysis has been expended to find the least expensive carrier-system design for short trunks. So far it seems that successful results have come in those cases where the telephone administration has requested the manufacturer to meet a particular set of conditions on certain types of cables (occasionally also on open-wire lines). The answer in those cases has been to provide a comparative price tag for each kilocycle of transmission spectrum versus cost of channel filters.

For the past 5 years, the Dutch telephone administration has been using a short-haul carrier system developed to its specifications. This was the result of a study conducted by Dutch and French telephone system development engineers.[3] They decided that the system should be usable on old carrier-operated cables as well as on newer types and that it should conform to CCITT recommendations with respect to quality. Double-side-band transmitted carrier as well as compandors were to be avoided. With these limitations, the approach was to save equipment cost by relaxing filter requirements.

Examination of the contribution of channel filter cost as a function of channel-frequency spacing led to a clear advantage for a carrier allocation in steps of 6 kc. Carrier terminal cost with this spacing is about one third of that with 4 kc according to these studies.

When such a system is applied to balanced or coaxial cable, its relative waste of frequencies will eventually wipe out the terminal equipment savings for circuits beyond a certain length. However, the study here shows that both these types of cables can be advantageously exploited, with their 50% excess frequency spectrum, up to lengths in excess of 60 miles. This is considerably beyond the main target for short-haul economy.

Similar systems, some using 8-kc spacing between channels, are now in use in most countries of Western Europe. The generally accepted prove-in point is at about 40 kilometers (25 miles) length. For shorter circuits, loaded voice-frequency cable pairs are still predominant.

A compromise with CCITT recommendations incorporated in these systems is their group allocation between the frequency limits of standard bands. This eases the transfer of such short-haul system groups in and out of larger capacity toll-carrier systems.

In Germany it is considered that these low-cost systems have made their contribution to encourage quick build-up of facilities by expanding transmission over old cables. But normally, growth of channel requirements between any two points is such that it does not pay to deal with 6- or 8-channel carrier systems. The 12-channel groups forming 120-channel cable or 960-channel coaxial systems are just about the lowest economical increment. On these facilities, frequency spectrum is considered too valuable to trade against savings in channel terminal costs.

It seems that there is no single correct answer to the question of what constitutes the cheapest short-haul carrier operation. Geography, importance of old plant investment versus new, local costs of equipment supply, projected-channel cross section: these are key factors in directing the telephone company's choice. At the present time in Western Europe, there are two schools of thought on this subject. One of these favors unconventional, low-cost equipment design, while the other favors the most efficient utilization of the transmission facility.

ADVANCES IN FILTER DESIGN

With outside system parameters generally standardized, European manufacturers' laboratories tend to concentrate on finding new tools to save money in carrier-equipment detail. Chief targets of attack have been the filters; not only channel filters, but also group filters since these are widely used in carrier-group transfer applications. Attempts to reduce filter costs can be classed in three categories:

1. Cost reduction in material and labor on filters of conventional design.
2. New design approaches to get the same filter results as before with fewer elements.
3. Active filters.

Conventional filter cost reduction has a fairly long history. The introduction of ferrites had its origin in Europe, no doubt stimulated by a drive to cut costs. Cheaper capacitors also are being used. The improvement of manufacturing techniques has allowed not only the complete replacement of the previously necessary mica capacitors by cheaper foil capacitors but in many places has also allowed the use of specially treated paper capacitors.

The improvement of the sintering procedures and the development of new ferrite materials allows the use of ferrite pot cores in favor of usually more expensive winding methods for conventional coils. The provision of inductance adjustments by slugs, films with magnetic material of varying thickness, or adjustable nonmagnetic shunts allows for final adjustment of the filters in the sealed enclosure and eliminates the job of capacitor padding.

The miniaturization of components allows for savings in cans, mounting, etc., and savings on over-all volume and is, as such, an independent factor in cost reduction.

One cost-saving design approach favored for some years has been the calculation of filter networks with equal inductance coils. Several calculatory transformations of the original design are performed until all coils in a given network are identical. However, this approach has lost ground with the introduction of the new adjustable ferrite pot cores.

More recently a heavy concentration on improved and cheaper designs of band-pass filters has taken place. The design according to the formulas published by Zobel[4] in 1923 has been more and more abandoned in favor of the design principles first laid down by Cauer[5] (1931). Cauer's design approach is complicated and difficult to apply. But recently a series of tables and monograms have been worked out which make it usable for routine filter problems.

Another significant step forward lies in the incorporation of the so-called "zig-zag" band filters. The special feature of these zig-zag filters is that with them coils can be saved at the expense of capacitors and vice versa. The final solutions, derived by a combination of these design methods, contain considerably fewer elements than conventional band filters. The additional big advantage is that they are easier to equalize; i.e., require less components in the equalizer.

Still another very promising approach under study is the provision of high attenuation peaks by equipment separate from the filter network. The peaks are achieved by separate magnetostrictive resonators. Such resonators can be made with Q's well in excess of 5,000. They considerably ease the design of band filters with extremely steep flanks. Carrier-frequency group-transfer filters, with separation bands of a few hundred cycles at pass-band frequencies in excess of 100 kc, is one of these cases.

Several leading manufacturers seem to put emphasis on the design of so-called active filters; i.e., filters where amplifying elements are incorporated to compensate for inherent losses. They permit such very lenient approaches to the over-all design that actually active resistance-capacitance filters can be used in carrier systems.

General Technological Advancements

The search for raw materials found at home is traditional in Central Europe where copper, nickel, and other basic materials have to be imported. The use of aluminum in wires and cables has made great progress.

Aluminum also is now widely used as cable sheath. Techniques have been developed to extrude the sheath over the cable, which provides a better fit than the previous methods. It is customary to cover the aluminum with a polyvinyl-chloride outer sheath which protects it from installation damage and corrosion.

Another European development has advanced the design of low-loss high-frequency coaxial cables. It uses a spiral-wound multilayer Styroflex insulation between the copper inner conductor and the aluminum sheath. It comes in a range of sizes up to several inches in diameter and can be advantageously used for radio and television station antenna transmission lines.

Styroflex has also found widespread application in capacitor manufacture as a preferred dielectric. It combines high dielectric constant with high insulation resistance. These capacitors have largely replaced the traditional mica types. European developments in cables and capacitors are finding increasing acceptance in the United States.

Still confined to relatively limited application is the metalized paper capacitor. This is a self-healing type of great interest in high-voltage applications. There is also a type of lacquer-dielectric capacitor. In this type the dielectric consists of a layer of lacquer baked on the metal foil. The objective here is to push size reduction to the limit. These are examples of the great ingenuity displayed by European engineers.

Customer Service Features Closely Associated with Transmission Development

While we are engaged in the introduction of direct-distance dialing, it is well to appreciate that this application of automatic telephony has a fairly long history in Europe. A large-scale regional direct-dialing network was set up in 1924 in Southern Germany. Similar developments have taken place in the Netherlands, Switzerland, and in Italy. Since the end of the war, progress in the countries of Western Europe has been rapid and consistent. European experience shows that direct-distance dialing acts as an incentive and thus raises the calling rate substantially.

There are other advantages which become apparent with use. Some telephone administrations have found that the fully automatic plant costs little more to keep busy on a 24-hour cycle than it does for the usual ten or so busy hours. Consequently, incentives have been set up to encourage subscribers to place their direct-dialing long-distance calls during the off hours. Rate reductions are very significant and the public has responded very favorably to this feature. Coin telephones suitable for direct-distance dialing have been developed. The customer has visual check of the number of coins he has inserted and realizes his time is up without intervention of an operator. He also has the option of predetermining the length of call by inserting the right number of coins at the start.

It has also been learned from experience that a direct-distance dialed call encountering a busy party presents a serious problem. An irritated subscriber repeatedly trying to make a connection can tie up a valuable long-distance trunk or several tandem trunks. One administration at least plans to overcome this difficulty by offering an optional "recall" feature. This arrangement will take over if the called line is busy, automatically make a connection within seconds after the line becomes free, and recall the calling party at the proper moment. It can be said that the advantages of automatic telephony are being utilized to a very high degree in Western Europe.

One other customer service approach which points up a fundamental difference between Europe and America should be mentioned here. That is the nonexistence of party lines as we know them in the United States. Some European telephone administrations have instead chosen the line-concentrator technique. A typical arrangement is for 3 lines to serve up to 16 subscribers. Usually, no more than 12 subscribers will actually be connected to such a concentrator. Full privacy is assured, and the telephone administration undertakes to allot these connections to subscribers with such calling habits that the resultant service is equivalent to single-party line. In Western Germany this technique has proved economical for subscriber-line lengths upwards of 1.5 miles.

On the whole, telephone service including long-distance facilities are well advanced in those countries of Europe which have their own development and manufacturing facilities, or which are closely collaborating with one of these leaders. However, the telephone density per population is still relatively low in many otherwise advanced countries. It seems that growth is paced by telephone administrations which struggle with political and financial difficulties while a very large segment of the population is either on record or otherwise known to be in the market for telephone service.

The traditional combination of mail, telegraph, and telephone often slows down telephone development. The postal service is considered a must, at any cost, and frequently is supported by income from the telegraph and telephone operations. Since telephony requires a relatively long-term investment in plant, money is difficult to raise where government administrations come and go every few months and where confidence is lacking. This condition is a marked contrast to the widely held notion that people have as many telephones as their cultural and economical status permits.

It appears rather that people as a whole are ahead of their own public administrations. Government-administered telephony is most advanced in a small number of countries which have long traditions of political stability and self-government (Switzerland, Nether-

Vol.
Pt. 1
1959

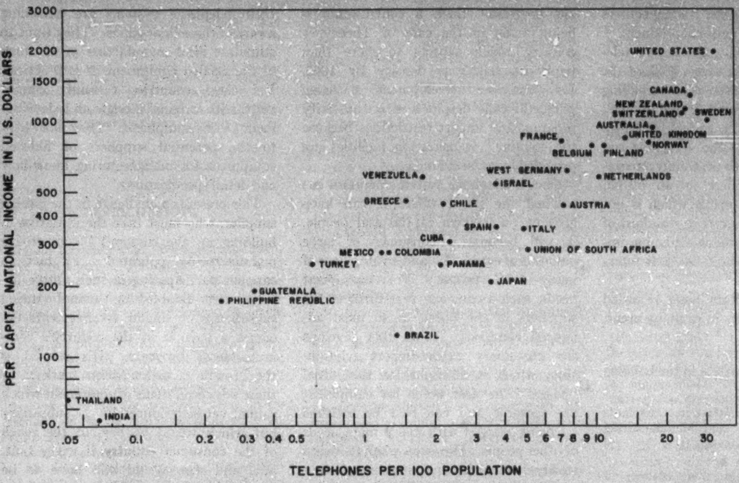


Fig. 4. Although telephone density generally corresponds to per capita income, political and cultural factors often cause considerable differences between countries of equal economic status

lands, Great Britain, and the Scandinavian countries).

Although the decisions on government versus private operation of telephone systems were made many years ago in most parts of the world, there remain a few countries where this is still a major issue. In Italy for example, a system of privately operating concessionaires has been in force since 1925. In these 30 years, the telephone density in Italy was raised ten times, reflecting the most spectacular progress of all countries in per-cent increase of subscribers. Also, local dial-telephone service is approaching 100%, and direct-distance dialing is already effective in larger areas than in many other parts of Western Europe. Now the government is considering taking over all telephone service and has called for a cancellation of the concessions, effective this year.

It will be interesting to watch how this issue is settled. The experience in Italy may have significance for many of the less developed countries around the globe which are now starting out with a system as weak or weaker than Italy's was in 1925.

Telephony in the Underdeveloped Areas

According to official reports, about 200 countries or independently administered areas have telephone service of one kind or another. But only about a dozen countries develop and manufacture telephone transmission systems. These same countries also operate about 85% of the world's telephones.[6]

What happens in the other 190 countries? Is it possible to classify them all as generally underdeveloped, and assume that telephone development is proportional to the standard of living? For a first statistical approach, telephone density and per-capita income have been plotted in Fig. 4. As can be seen, the general trend is unmistakable, particularly in the high-income countries. There is a high positive correlation between per-capita income and telephone usage.[8]

But a closer study indicates that such generalizations do not necessarily account for the position of less-developed countries. For example, Greece and Austria have about the same per-capita income, yet Austria leads Greece in per-capita telephones by a ratio of over four to one. The ratio is 10 to 1 for Japan versus Guatemala, two other countries with about equal per-capita income. Different people have different uses for the same amount of income to spend. Telephone service is more attractive when more people with telephones are accessible to a subscriber. Thus it is a service that feeds on itself. If the local telephone company shows initiative in setting up offices and lines for a growing number of connections, then people will be induced to participate. If, however, there is individual distrust in the quality of public services, people, regardless of income and social stature, will rely on other forms of message exchange, or will remain more isolated from each other.

Telephone density figures alone are not indicators of a whole country's telephone development. For example, the Union of South Africa and Italy have, at the present time, the same density of about 6 in 100 population. But in South Africa, due to the political status of the natives, only about 20% of the population have an opportunity to participate in telephone development. Thus, the true telephone density in that country may be closer to 20 in 100 of the white population. It could then be concluded that South Africa has a more highly developed telephone system than many European countries.

This example also points out that telephone development may be judged in relation to factors such as business activity, rather than population and per-capita income figures. This is particularly true in the early phases of development. There is no country so backward or so disinterested in communication that it could not justify from 2 to 10% telephone density in its big cities. The cities of the Orient are teeming with small and large businesses which exchange messages among themselves, and with many of their customers. In addition, communication between the city and the food-producing countryside is an absolutely fundamental need. The cities must also utilize communications to function as the trading centers for the foreign sale of national products.

From this viewpoint, the lag in telephone development in about 90% of the world's countries is even more appalling than the usual figures would reveal. While it is true that large-scale build-up of communications in the very long run must be in step with a country's economic well-being, there is an initial, basic development possible which is not dependent upon the average standard of living. This development finds its impetus in the big cities and in their interconnecting networks.

Two principal factors have retarded this phase of telephony in so many areas. They are:

1. Lack of native initiative in the building of telephone networks.

2. Lack of outside initiative in transplanting foreign know-how and manufacturing ability.

Native initiative is, of course, a function of economic and political stability. But it is also necessary for a country's political administration to recognize that communications development cannot be turned on and off at will. Where this recognition has been gained, there is usally a nucleus of people for whom telephone development is a lifetime job. These people will develop a plan which requires, among other things, the availability of capital. At this point the program must fight for survival with all the other more or less worthy proposals for spending the public's money. Since the telephone people can promise neither guns nor butter, nor tangible profits for any major pressure group, they do not usually fare very well.

Given time and a few stable years, the telephone system may be substantial enough to grow modestly on its own earnings, at which time its problem will shift to one of preventing its earnings from being siphoned off for use elsewhere. But the most crucial growth period for a telephone system is precisely during its infancy, when it must draw on outside capital. It is like denying milk to a helpless baby, while a starving adult would fight for his own survival.

To spur initial telephone development in the baby phase, outside capital may be brought in, provided there is a native nucleus with a plan and a will to succeed. Where governments manage the telephones, this financing may come in the interest of military necessities such as the North Atlantic Treaty Organization communications program or perhaps as part of a United States foreign-aid program. Such a contribution is being made in the case of Turkey, a country which expects to more than triple its telephone density by 1965. Its subscriber development is being pursued locally despite a severe austerity program and import restrictions because the country's transmission facilities are provided by outside funds.

Once a telephone system operation has reached the point where it can keep growing on its own capital and people, it will become a consumer of large amounts of equipment and an employer of many local people. For equipment needs, such a company first turns to the suppliers in the dozen or so most advanced countries. From this develops the customary export-import relationship, which is all right for the initial phases. The first needs for equipment are sporadic and can best be satisfied by being merged with many more needs of other people. However, when there is a steady demand for equipment in sufficient volume to support a basic manufacturing organization, steps should be taken to develop native manufacturing.

Continued dependency on import can become a major obstacle to a telephone company's growth for the following reasons:

1. The purchasing telephone administration must set up a detailed specification for the foreign supplier to meet. This may or may not be considered necessary by the supplier, but it is always being done to protect the purchaser. Since it is likely that the supplier knows more about the equipment and its applications, this specification becomes an obsession for its authors and calls for unreasonable ingenuity, investigation of marginally important points, and procrastination in tying down the requirements.

2. It is one thing to get an appropriation for telephone expansion in the first place, but quite another to get an import license for equipment and, finally, an allotment of the necessary foreign currency. Telephone organizations waiting for equipment subject to these obstacles may have their plans set back by several years.

3. The cost of equipment is usually higher than it would be if native labor could be efficiently employed.

4. There is no interchange of personnel and little advance knowledge of equipment features as there always is where manufacturing and operating technical people work in geographic proximity and speak the same language.

5. An operation which must depend entirely upon foreign-made material can never become very popular. People do not like to think that their telephone bill makes money for some remote foreigner's pockets.

More and more countries whose administrations are anxious to advance their telephone systems are becoming aware of these limitations. They want to stimulate local manufacture of as much of the needed equipment as is practical. Yet these countries certainly cannot start with original design and development of this equipment. They must look to the potential suppliers of finished equipment for manufacturing know-how and detail specifications.

This presents a challenge to the foreign supplier, who must take the initiative in building up the customer's competitive manufacturing potential. In fact, a supplier participating in such native development establishing manufacturing subsidiaries or similar arrangements becomes a partner in the country's communications program. His reward is the growth of a long-term market for those advanced items of equipment which cannot yet be produced by the subsidiary but which would be beyond the reach of the consumer country if every bolt, nut, and wire would still have to be imported.

There are signs that this approach is finding increased favor. Starting with modest partial assembly operations and progressing gradually towards more intricate production work, telephone communications equipment is now being made in such widely separated countries as Mexico and India, to mention only two. These factories turn out more than needed equipment: They raise a crop of native engineers, technicians, and skilled workers. They contribute directly in making an entire country a better partner in international communications. They help immeasurably to overcome that country's inferiority complex, hence its animosity towards imported technical contraptions.

Ultimately, each country with a multimillion population, and with language, customs, and administration distinct from its neighbors, will have to have a communications manufacturing industry if it wants to have communications. The process will be accelerated in those countries where native initiative is a favorable incentive for the needed foreign investment.

References

1. Das Neuve Kunststoff-isolierte Fernkabel der Deutschen Bundespost, H. Duell, W. Wild. *Jahrbuch des Elektrischen Fernmeldewesens*, Windsheim, Germany. 1954–1955.

2. General Interconnection Plan 1954–1958. Comité Consultatif International Telephonique, Geneva, Switzerland, 1954.

3. Système Simplifié à Courants Porteurs Pour Courtes Distances, J. Hurault, G. Bast. *Cables et Transmission*, Paris, France, no. 3, 1953, pp. 185–217.

Vol. 7
Pt. 1
1959

4. Theory and Design of Uniform and Composite Electric Wave Filters, O. Zobel. *Bell System Technical Journal*, New York, N. Y., 1923, pp. 1–46.

5. Theorie der Linearen Wechselstromschaltungen (book), W. Cauer. Academische Verlagsgesellschaft, Leipzig, Germany, 1941.

6. The World's Telephones. American Telephone and Telegraph Company, New York, N. Y., 1956.

Discussion

Roman Krzyczkowski (Westinghouse Electric Corporation, Baltimore, Md.): This comprehensive paper by Mr. Wallenstein "is not intended to be a complete list" of important trends and developments, hence I should like to concentrate only on the most important problem which, in this discusser's opinion, did not obtain full justice in the otherwise excellent review.

The problem revolves around the question of demand for service: How many circuits must be provided and at what (uninterrupted) distances? Is television transmission involved?

Mr. Wallenstein gives a good description of multipaired cables for working 60 and even 120 channels per pair. However, a look at the map of principal European networks shows that these are still largely in experimental stages and the great majority of high-traffic routes are served by standard coaxial cables. This is, at least partly, a result of the fact that while two separate cables are required for 4-wire working (due to crosstalk difficulties), a single coaxial cable can easily give equivalent capacity. On first costs alone, if the price of 2-tube coaxial is taken as 100, two 4-quad cables capable of taking 60 channels per pair would cost at least 140, and two 7-quad cables, 170. Moreover, the cost of terminal equipments is about the same for the capacities in question and that of line repeaters is actually higher for multipaired installations. Finally there is an almost universal demand for television channels capacity that can be provided in cable only by coaxial techniques.

Extensive studies made show that for capacities of some 300 telephone channels and more, and for medium and long hauls (above, say, 60 miles), the coaxial techniques offer the following advantages over multipaired cables:

1. One cable is required instead of two.
2. Hence the first and annual cable costs are smaller.
3. Duct charge is also smaller.
4. Installation is cheaper and quicker.
5. Transport costs also are lower.
6. Foreign import duties, whether levied per value or per mile of cable, are lower.
7. Quantities of metals are generally smaller, and hence better hedging results against base price fluctuations.
8. Smaller stocks of metals must be carried.
9. Only coaxial cable ensures a possibility of television transmission.

There are also other basic forces at play. Designers of 60- or 120-channel cables fight against natural laws since crosstalk difficulties increase quickly with frequency. But in a coaxial cable the exact converse is true: the higher the frequency the less crosstalk there is to worry about. Thus, existing coaxial cables can easily accept 12-megacycle/second bandwidth. Hence studies are being made to halve existing repeater spacings (from 6 to 3 miles) and thereby to increase threefold capacities of existing installations (from 960 to 2,880 telephone channels) per two coaxial tubes.

It is clear that the cost of fitting 12-megacycle/second equipment on existing cables is only a fraction of a new installation and hence this technique may prove very advantageous to financially harassed administrations.

In specific conditions coaxial techniques can be proved in distances of some 20 miles and capacities of no more than 60 telephone channels.

It may be deduced from the foregoing that the future of European networks will be significantly shaped by competition between coaxial cables and radio links, with ingenious multipaired cables filling only very special requirements.

Coming to other points, it is perfectly true that CCITT recommendations make it difficult to use such techniques as compandors or 6-kc spacings between adjacent telephone channels. On these points Europeans quite often envy American practices. However, once existing networks came into being, any systems incorporating these devices can offer now only very marginal advantages. Careful studies show that not much can be gained from relaxation of CCITT recommendations; especially since the change from 4 kc to 6 kc is not advantageous when the whole system is considered, i.e., a cheaper filter is balanced against wasted space in (expensive) cable.

At least in one country party-line arrangement is well known. In Great Britain 2-party subscribers form a large percentage of the total. On the other hand, concentrators there are only under study.

The close interrelation between telephones and mail has many advantages for the former. Mail normally more than pays its way and can quite often subsidize underdeveloped telephone networks working below their potential capacities. On the other hand, excellent postal arrangements (several deliveries daily, same-day delivery within metropolitan areas of mail posted before 4:00 p.m., etc.) and their low cost are serious competitors to telephone traffic, especially to long-distance calls.

Finally, with regard to telephony in the underdeveloped areas: All major countries in Europe (Great Britain, France, Germany, Italy, Sweden, Belgium, Holland, Switzerland) have their own fully developed telephone industries. Names like Automatic Telephone & Electric Company, Ltd., Siemens, Ericsson, and Phillipps, denote what is best in the telephony field. Also, these and other countries have many plants, some of which are subsidiaries of either American principals (such as International Telephone and Telegraph Corporation or Autelco) or of the foregoing European companies. Both groups of interests foster most actively development of local manufacture. Out of many examples let me quote one: the ultramodern Indian Telephone and Telegraph Company into which the British Automatic Telephone & Electric Company, Ltd., put the best of its knowledge, experience, and good will.

The principal factor which retards the development of telephony in many areas is simply the lack of capital; neither local nor outside initiative are lacking. Suffice it to say, that in many European companies exports (including the export of know-how) exceed 50% of their total turn-overs. What is wanted is peace and an increased standing of living. As the American experience shows, given these two factors, the telephone industry can become the most prosperous undertaking and the quality of telephone service will be unequaled by any standards.

The discusser trusts that these remarks may be found of interest by students of telephony problems in countries other than America as a supplement to the valuable paper of Mr. Wallenstein, who deserves to be congratulated for his initiative in originating this discussion.

G. D. Wallenstein: Mr. Krzyczkowski's discussion of the merits of coaxial over balanced-pair cable transmission could well be the subject of another paper. The stated purpose of the paper was to highlight those European developments which are *different* from United States and Canadian practice. Coaxial systems closely similar to those used in Europe have been fully developed here by the Bell system and have been adequately described in several AIEE papers.[1–3] By contrast, no further development of all-carrier frequency balanced cables has occurred here since the days of the Western Electric *K* System. The improved European multipair systems form the backbone of telephone networks in a good part of Western Europe where they have long left the experimental stage. The official CCIF "General Interconnection Plan 1954–1958" lists many thousands of trunks of this type in existence or to be provided.

Since the presentation of this paper in 1957, there has been one new development of a small-gage coaxial system in England. This type of system, of 300-channel capacity, will bring coaxial techniques economically more in line with balanced-pair cables in applications for smaller circuit groups.

The belief that the mails more than pay their way is not borne out in leading countries of Western Europe such as France and Germany[4,5]. In the United States, difficulties of balancing the postal budget have been well publicized. The point is well taken that some types of postal communication service compete favorably with the telephone. Generally it would be advantageous to see to it that such competition be carried on under different managements than under the same monopoly.

Some British companies have indeed made a good effort in India to assist in the growth of local development and manufacture. In good part this is due to the initiative of the Indian government which places telecommunications equipment manufacture under national control. The point of my paper is to stress that more creative native participation is needed instead of outright export. This approach has not been necessarily considered good business by export-minded manufacturers. The often quoted lack of capital is only a symptom for a lack of initiative wherever there exists capital hoarded and otherwise idle. This is true for many underdeveloped areas and represents a latent immediate potential which could be tapped long in advance of broad-scale improvements in standard of living.

The comments received certainly indicate that this subject could benefit from more exhaustive studies.

REFERENCES

1. WIDE BAND TRANSMISSION OVER COAXIAL LINES, Lloyd Espenschied, M. E. Strieby. *AIEE Transactions (Electrical Engineering)*, vol. 53, Oct. 1934, pp. 1371–80.

2. PROGRESS IN COAXIAL TELEPHONE AND TELEVISION SYSTEMS, L. G. Abraham. *AIEE Transactions*, vol. 67, pt. II, 1948, pp. 1520–27.

3. THE L3 COAXIAL SYSTEM, C. H. Elmendorf, R. D. Ehrbar, R. H. Klie, A. J. Grossman. *Ibid.*, vol. 72, pt. I, Sept. 1953, pp. 395–418.

4. TELEPHONE SERVICE MADE £5,3 MILLION SURPLUS. *Post Office Telecommunications Journal*, London, England, Spring, 1959.

5. POSTAL COMPARISONS. *The Economist*, London, England, July 5, 1958.

The Present Status and Anticipated Progress in the Field of Insulating Materials

THOMAS D. CALLINAN
NONMEMBER AIEE

INSULATING MATERIALS have become increasingly the subject of serious study within recent years. Seven books[1–7] and at least ten full-scale reviews[8–17] on insulating materials have appeared within the last 10 years. These do not include a monthly trade paper, *Insulation*, dedicated wholly to the subject,[18] or the annual *Digest of the Literature on Dielectrics*, published by the Conference on Insulation of the National Research Council.[19]

In addition, *Electrical Engineering*, *Electrical Manufacturing*, *Transactions* of the Institute of Radio Engineers (Professional Group on Component Parts), and the Society of Plastic Engineers present, as a continuing policy, articles on electrical insulation and its uses. Furthermore, the theoretical aspects of dielectrics appear frequently in such accredited organs as the *Journal* of the American Ceramic Society, the *Journal* of the American Chemical Society, the *Journal of Applied Physics*, the *Journal of Chemical Physics*, the *Journal of Physical Chemistry*, *Physical Reviews*, and *Reviews of Modern Physics*. Articles on one or more aspects of dielectrics appear also in European and Asian journals devoted to chemistry, engineering, or physics.

The increasing interest shown in dielectrics as exemplified by publications has been accompanied by a startling growth curve for talks and papers delivered on the subject at annual meetings of learned societies. Both the AIEE and the Institute of Radio Engineers sponsor active programs on insulating materials and their use and evaluation in components. Attendance at symposia has been as high as 400 members and guests. The number of papers per session is usually between four and six, and it is common at the AIEE annual winter meeting to have four to six sessions devoted to insulation topics.

The National Research Council, with a membership of 400 to 500, meets annually to hear about and to discuss electrical insulation. The Electrochemical Society sponsors an annual 3-day symposium on electrical insulation and has a permanent group of 100 members enrolled in electrical insulation division. Meetings of the National Research Council and the electrical insulation division, neither of which require formal written papers, have become so packed with informal talks, disclosures of new materials, and discussions that *Electrical Manufacturing* devoted ten editorial pages to report last year's happenings.

Besides the established society meetings, 1- to 3-day sessions on the subject are being sponsored increasingly by government organizations and universities. This amazing growth of interest in a subject as specialized as electrical insulation bespeaks intellectual problems which numbers of dedicated and well-trained scientists find worthy of attack.

The books which have appeared present the viewpoints of physical chemists (Smyth), chemical physicists (Bottcher), engineering physicists (von Hippel), mathematical physicists (H. Frohlich), and engineers (Whitehead).[1–5] The fact that men of such seemingly diverse backgrounds find themselves in the same field of study arises from the nature of dielectrics considered as an intellectual discipline. It may be studied as a section of chemistry or engineering or physics, without regard to the other sciences, but the subject of dielectrics has depth only when illumined by the laws of all three. Its composite nature may be visualized by Fig. 1.

To the electrical engineer, insulation has always had a functional aspect which has impelled him to make contributions to the subject, not particularly derivable from the more fundamental approaches of the other disciplines. The engineer, interested in dielectrics, has done far more than just adapt available materials to his equipment. His laminated structures, which possess high dielectric strength, have not yet been derived from theoretical foundations. Again, his use of liquid-impregnated solids, while understood empirically, has yet to be quantitatively developed.

To the physicist, the nature of dielectric polarization has proved an intellectual forte and has resulted in predicating theories of energy storage and energy transport in insulating materials. Incidentally, the basic problem

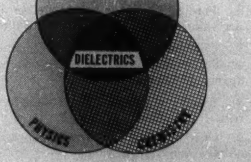


Fig. 1. The composite nature of dielectrics

Paper 58-1208, recommended by the AIEE Electrical Insulation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958 and represented at the Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted June 9, 1958; made available for printing August 15, 1958.

THOMAS D. CALLINAN is with the International Business Machines Corporation, Poughkeepsie, N.Y.

Vol.
Pt. 1
195

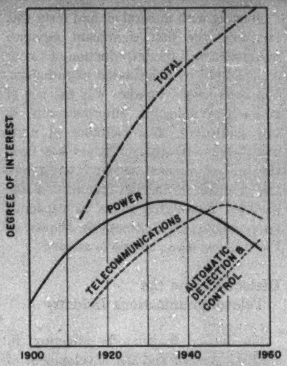


Fig. 2. The contribution of electric power, telecommunications, and automatic detection and control to dielectric studies

of relating the macroscopic properties to molecular ones was the subject of at least 20 published articles last year, and the total number on the subject since the century-old work of Classius and Mossotti is now well over 1,000.

To the chemist, dielectric phenomena have served both as a tool for the elucidation of molecular structure and as an incentive for developing new and improved gases, liquids, and solids. The chemist has found the electrical industry an apt place for pursuing his tendencies toward developing and molecularly engineering new compounds of greater moisture resistance. He also has found that his interest in uniformly fine inorganic structures has been encouraged by the electrical industry and has resulted in the development of not only the high-voltage ceramic insulators, but the magnetic ceramics as well.

Dielectrics has indeed been a fruitful field for all three disciplines and each recognizes the contributions of the others to it. An examination of the forces which have led to the interest in insulating materials and the forces extant today which are motivating further research should permit a critical evaluation of the present state of dielectrics and anticipated progress in the field of insulating materials.

Sources of Industrial Interest in Dielectrics

The development of insulating materials into the subindustry of today has arisen from three enormous forces, namely 1. the electric power industry, 2. the telecommunications industry, and 3. the new industry concerned with automatic detection and control which includes recognition and guidance systems, detectors, memories, computers, and servomechanisms.

The electric power industry had, as its basic proposition, the idea that the generation and transmission of electric power was economically feasible and that the availability of cheap power would result in its exploitation. It is rather difficult today to comprehend the anxiety, fear, and worry which passed, at one time or another, through the minds of the founders of the electric power industry. Today the problems of the industry are ones of finance. Scientific meetings deal with such problems as reliability, life expectancy, and improved distribution techniques, all of which become important only when one is extremely successful.

The telecommunications industry, having as its purpose the reliable transmission of intelligence, has been the second fruitful source of employment for insulation specialists. This force grew, beginning in the second quarter of the 20th century, into all that is now known as electronics. It involved the subindustries of telephony, radio, television, radar, high fidelity, and the movies. While the basic mechanism or driver was the vacuum tube, the insulation specialists found and solved problems in high-frequency cables, capacitors, and hook-up wires.

Automatic detection and control is the third and latest force to motivate dielectricians. It differs from the work of the previous two periods in this respect; rather than adopting materials to sustain high voltages or to pass high-frequency currents without loss, it entails utilizing dielectric phenomena to identify an event. Such an event is compared with one stored in a memory and, depending on the outcome of the comparison, a machine is made to act in accordance with a set of preissued instructions. It would be an understatement to say that this problem is an intellectual challenge equal to that of the previous two forces. It will certainly require an interaction of many disciplines to solve these latest problems.

To suggest the contribution of each force toward developing new insulating materials over the past 58 years, Fig. 2 pictures the general interest shown by investigators through their patents and publications. The figure suggests that there is a strong but declining interest in developing new insulating materials for the power industry at this time and even a slight drop in interest in classical communication problems. But these are specialized views and a decline in interest is to be expected as more and more problems in the field are solved. The general interest in insulating materials has increased over the years as shown by the dotted line. Incidentally, if instead of patents and publications the expenditures of the electrical industry and its suppliers are taken as a measure of the force for developing new insulating materials, there is no observable decline of interest in any one of the sectors.

Dielectrics and the Electric Power Industry

The chief problems of the electrical industry in the first days of its establishment required the development of high-voltage transformers and cables, large generators, and subsidiary equipment such as reactors, power factor correction capacitors, lightning arresters, and circuit breakers.

Some of the problems were solved quickly and satisfactorily. Thus, porcelain bushings, once accepted, have not been replaced. Again, the lead-oxide pellets of the lightning arrester are standardized. On the other hand, the demand for longer life in more compact equipment has led to a continuing demand for new transformers and generators, motors, and high-voltage power-frequency capacitors.

From this viewpoint there were four distinct kinds of insulating materials required: 1. solids for slot armour, insulating cylinders, and turn insulation, 2. enamels for wire, 3. thin separators for capacitors, and 4. liquids as impregnants and coolants.

STRUCTURAL INSULANTS

The high-voltage transformers and the heavy generators and motors required insulation which could resist the severe

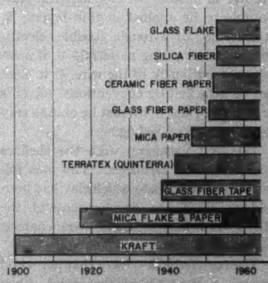


Fig. 3. Structural insulants developed because of electric power needs

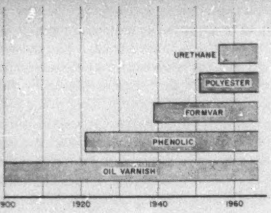


Fig. 4. Wire enamels developed because of electric power needs

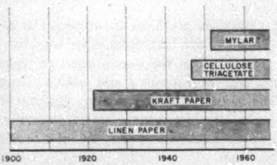


Fig. 5. Capacitor spacers developed because of electric power needs

mechanical and thermal shocks to which the equipment was subjected, which was inert towards moisture, and which had reasonable dielectric strength and moderately low electrical losses. Because heavy electric equipment is prone to be subjected to serious temperature and mechanical overloads, designers of such equipment have demanded that the insulation used be thermally stable. As a result of this demand there has been developed products like terratex and quinterra, alsifilm, glass fiber yarn, glass fiber paper, silica fiber paper, mica paper, synthetic mica paper, and glass flake paper (Fig. 3).

WIRE ENAMELS

The requirements of a good wire enamel are five in number: 1. the material must be able to be applied in thin layers; 2. it must dry to a tough, flexible, electrically nonconducting material; 3. it must adhere tenaciously to the wire, 4. it must not be attacked by conventional solvents or moisture, and 5. it must be able to stay flexible after thermal aging.

The first varnishes were the shellacs and linseed oil varnishes. This was followed by varnishes to which clay had been added. With the development of the alkyd varnishes, synthetic wire enamels became available.

Then came the discovery that polyvinyl formal, a clear transparent solid, when mixed with phenol became plasticized. This mixture in turn could plasticize a phenolic varnish. The enormous success of Formvar wire enamels is deserved.

There have been introduced lately a number of new enamelled wires having thermal stabilities superior to those of Formvar enamel. This historical development of these enamels is shown in Fig. 4.

HIGH-VOLTAGE POWER CAPACITORS

Originally capacitors were of two kinds, air and mica.

The development of the armature or rolled-type capacitor transformed the manufacture of capacitors into a high-speed process. Such a unit had two possibilities for quality improvement which appealed to the chemist; he was able to change the chemistry and thickness of the spacer and he was able to change the chemistry of the impregnant. As a result, over the period of the last 50 years, there have been many worthy improvements in the life, thermal stability, and size of such units.

While paper is still employed as the dielectric, a significant change in the chemistry has occurred. Early theories suggested that fine linen papers should be used, and much work was expended in cleansing the stock; now it is known that select impurities in kraft pulp actually extends the life of liquid-filled capacitors.

Attempts to replace the paper structure with continuous polymer films has had some success. Considerable work was done on adapting cellulose triacetate to high-voltage capacitors, but it was not until the highly uniform polyethylene terephthalate was produced that a significant advance was observed. Furthermore, it has been found possible to make paper uniformly in thicknesses of 0.0002 inch, a reduction of 60%. The uniformity of the paper has been improved enormously in the meantime. These improvements are pictured graphically in Figs. 5 and 6.

LIQUID IMPREGNANTS and COOLANTS

Liquids are employed by designers of electric power equipment to serve as coolants, or to confer dielectric strength, or to store energy. Sometimes the liquid serves multipurposes as shown in Fig. 7. Over the years liquids have been developed having both satisfactory coolant and dielectric properties. They range in boiling point from the Freons up [24.6 C (degrees centigrade)] through trichlorobenzene to the silicones (> 200 C). Molecular engineering of liquids with greater dielectric strength has not been accomplished, but the synthesis of designed molecules having high-energy storage properties has been very fruitful.

Starting with mineral oil and rosin wax mixtures, the first significant advance was made with the introduction of castor oil, followed by chlorinated naphthalene. The next major advance was the use of pentachlorodiphenyl. Improvements in the uniformity of capacitance of units over large temperature ranges has been the subject of more recent interest and has resulted in the development of the partially fluorinated esters as capacitor impregnants. The trend is shown in Fig. 8.

Dielectrics and the Telecommunications Industry

The second major force affecting dielectric research and the development of new insulants has been the demand for the reliable transmission of intelligence. This became increasingly a subject for study and development by dielectrics experts in the second quarter of this century. The phenomenon, while still that of energy transport, was no longer limited to 60 cycles per second in frequency. In terms of insulants, there

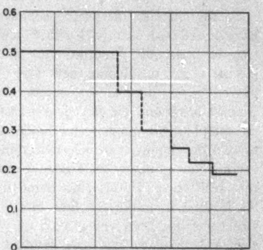


Fig. 6. The reduction of capacitor tissue thickness developed because of electric power needs

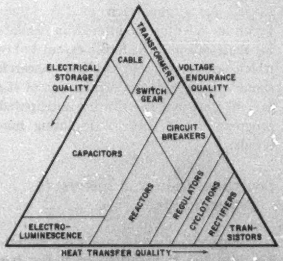


Fig. 7. The properties of liquids in relation to their uses by the electric power industry

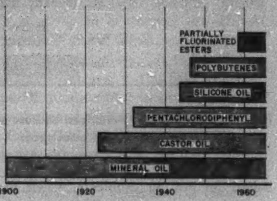


Fig. 8. Capacitor impregnants developed because of electric power needs

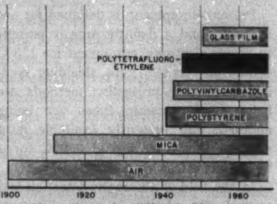


Fig. 9. Radio-frequency capacitor dielectrics developed because of telecommunication needs

l. 78
1
59

were requirements for improved materials for radio-frequency (r-f) and d-c capacitors, r-f transmission cables, and radomes.

Radio-Frequency Capacitors

The communications industry in its first days of growth was interested, so far as capacitors were concerned, with the transmission of high-frequency currents of low voltage. There existed from the beginning, the pedagogue's dream, a capacitor with air as the dielectric which was practically lossless over the frequency and temperature spectrum.

Based on the familiar equation relating capacitance of a parallel plate capacitor directly to area of the plate, and inversely to their separation, mechanical engineers had made strenuous efforts to build air capacitors of high capacitance by bringing the electrodes as close together as possible. Innumerable geometries were proposed to take advantage of this effect, but ultimately it became realized that the dielectric strength of air limited further developments.

While there was a certain flavor of prosperity in large air capacitors, Curtis' classic work on mica capacitors clearly demonstrated the possibility and value of miniaturizing.

With World War II came two problems which required enormous effort to solve: 1. the unavailability of mica suitable for capacitors and 2. the need for high-voltage d-c units with compact efficiency resulted in the development and adaptation of low-loss organic and inorganic materials for capacitors. Thus, polystyrene, polyvinyl carbazole, and polytetrafluorethylene (organic polymers having no permanent electric moments) were fabricated into high-frequency units for band-pass applications. Organic liquids such as mineral oil, silicone oil, and dibutyl sebacate were used in conjunction with parallel plates for resonant tank circuit uses. Meantime, strenuous efforts were made to synthesize bodies which could be fashioned into flakes or continuous sheets. Among the materials which emanated from this work were the fluorophlogopites (synthetic micas) and the low-loss glasses. The results of this effort, therefore, were eminently successful and a gamut of mica substitutes are now available for high-frequency work. This is shown schematically in Fig. 9.

Direct Current

At first the demand for the high-voltage d-c oil-filled unit having nonflammable properties seemed to involve nothing but an adaptation of the a-c pentachlorodiphenyl unit. To everyone's chagrin, it was soon found that this was not true; the units which have life expectancies of 14 to 30 years on alternating current were failing in periods of one week on direct current even when severely derated. The solution came with the establishment of the value of additives such as anthraquinone. Today it is a recognized practice to introduce such additives in insulants (both liquids and solids) for both d-c and a-c applications.

There has always been an enormous interest in low-voltage high-capacitance units having life expectancies as great as those of other components. The electrolytic capacitor, while fabulously successful in storing energy, has been seriously limited by short life times and low operational temperatures. The tantalytic capacitor has made a heartening advance recently in a field in which none was reported in a very long time. For example, there is now available a nonacid electrolytic unit capable of withstanding 125 C for 2,000 hours with a drop in capacitance of only 12%.

Radio Frequency

Just as polar molecules in capacitors caused undue losses to develop, so too the conventional wire and cable coatings were lossy and attenuated a high-frequency signal sent down the transmission line. The demand became eloquent when radar became a necessity and resulted in the rapid wartime development of polyethylene insulated cable. A second such cable employed polystyrene and even polytetrafluorethylene has been used on special projects. Wire and cable coated with polyethylene, which has been cross-linked by ionizing radiation, has also entered the market. The general growth of these items is shown in Fig. 10.

Radomes

Another problem of intelligence transmission which intrigued the dielectrics expert was that of radomes. Here the question was: "How can an antenna transmit and receive radio waves when it is housed inside a protective dome or canopy such as the nose cone or belly of an airplane?" The material must be mechanically strong and not attenuate the signal. As shown in Fig. 11, the materials having the best mechanical properties were ones which underwent maximum dielectric absorption in the r-f range. This led to the serious consideration of dielectric mixtures. In one case glass fiber cloth, the mechanical

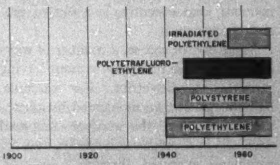


Fig. 10. Radio-frequency cable dielectrics developed because of telecommunication needs

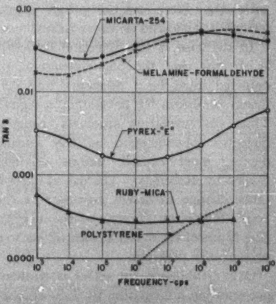


Fig. 11. The dielectric loss factors versus frequency of certain structural plastics

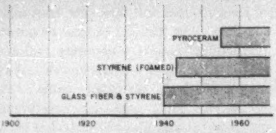


Fig. 12. Radio-frequency radome dielectrics developed because of telecommunication needs

strengthening ingredient, was impregnated with polystyrene, the low-loss dielectric, to yield a canopy. For lower losses, but at the cost of strength, polystyrene foam was used. The most significant advance has been made only recently, however. The material, pyroceram, is an inorganic body of high mechanical and thermal stability and having low electrical losses. The advances are shown in Fig. 12.

Dielectrics and the Automatic Detection and Control Industry

The third force which is making itself felt increasingly may be called automatic detection and control. This entails the development of equipment capable of recognizing a select phenomenon or condition, comparing it with observations previously stored in a memory, and executing a series of prestored commands.

There are, of course, a number of ways of solving these problems without having recourse to dielectrics. For example, light pulses can be monitored by photomultiplier tubes, the number compared with one in a magnetic memory, and the command executed through a conventional servomechanism. Dielectrics enter when a phenomenon associated with the polarization of insulators can be used in one of the three stages. Effects which are being studied include those associated with the Kerr electro-optical effect, ferroelectricity, and electrostriction.

In the case of the detection device it is obvious that the phenomenon to be detected must be a necessary and sufficient condition for developing a signal. Thus, the insulation specialist, using a dielectric phenomenon as the sensing device, must exploit unique changes caused by ambients. The use of capacitance change to detect crevasses at the South Pole is an example of the adaptation of one of the oldest known effects to a present-day problem.

Once established, the search instrument presents the information to that segment of the automatic detection and control instrument called the memory. The basic requirement of any memory is that it be a substance which undergoes a nonlinear change when subjected to a change in ambient.

To the dielectric expert, the switching of a barium titanate crystal from one state of polarization to another by the action of an electric field is an example of how insultants may play a role in automatic detection and control. This is one of the reasons for the widespread interest in the ferroelectrics.

The third problem associated with automatic detection and control involves the execution of a command. The physical movement of charged particles as in xerography, or the elongation of a piece of matter through the action of a electric field, are examples of third stage of automatic detection and control.

While phenomena such as the Kerr electro-optical effect were observed in the last century, the forces motivating research in this field have been sporadic until recently. The work on ferroelectrics is of relatively recent origin even though the piezoelectrics have been studied for a century. In terms of automatic detection and control, Fig. 13 shows the early state of development which now exists.

The Present Status of Dielectric Materials

In summary then, it would appear that at this time there are available a plethora of materials which have been developed to satisfy the requirements of the power industry. While there are certainly improvements to be made, there are available, for use by an engineer, a host of materials from which he may choose. He has a wide range of liquids, solids, and gases available, because in many cases they possess qualities which make them useful in fields other than the electric power industry. Actually, because of this widespread availability, the insulation user finds himself more and more engaged in functionally evaluating the items and from his data deciding which one best suits his needs.

In communications, the atmosphere has changed significantly from the inception of this industry. There are here also an enormous number of materials suggested by their manufacturers for use. Actually, the tremendous nondielectric interest in such things as polyethylene, polystyrene, and teflon have made available to the practicing engineer a continually improving product which he can adapt to his purposes. On the other hand, he is experiencing difficulty in obtaining dielectric materials suited for use in ambients which are certainly not normal. These are the requirements of very high temperatures and exotic radiation environments where telemetry and communications are necessary.

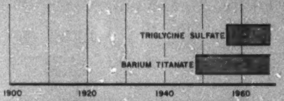


Fig. 13. Ferroelectrics developed because of automatic detection and control needs

Present equipment is designed for use at 85 C, and a definite urge is present to design electronic equipment for 125 C operation. As stop gaps, shielding, cooling, and other protection methods are being used, ceramic tubes are being adapted, and presently available data suggests that their use at 300 C for a short time is feasible now.

In the field of automatic detection and control, the picture is quite different; there are very few materials available to the design engineer, and these are not completely satisfactory. As for such niceties as reliability and stability in exotic environments, these are only in the planning stage.

Anticipated Progress in Insulating Materials

From this analysis an answer to the question of anticipated progress in the field of dielectrics can probably be extrapolated.

The field of power transmission is deeply committed to elucidating the laws of decay of insulation and to determining effects of ambients other than heat on the life expectancy of the items. As a result of such study, the evaluation of old as well as new insulants will become a matter of science. It also seems inevitable that there will appear sporadically new thermally stable enamels, solid insulants, both flexible and rigid, and even new liquids to act as coolants and energy storers.

The revitalization of interest in electrolytic capacitors and their elucidation in terms of solid-state physics will undoubtedly lead to the preparation of new miniaturized capacitors. The capacitance which arises at a *p-n* junction is one such phenomenon likely to be exploited.

The problem in telecommunications which has most significance to dielectric specialists is that of the low electric strength of solids and liquids at high frequencies. In many cases the strength of a gas at radio frequencies is greater than that of the best plastics. Unlike the situation in evaluating the life expectancy of power insulants, there are no theoretical guideposts. It is hoped that the future progress in this field will come from a fundamental knowledge of the mechanism of electric breakdown.

In the field of automatic detection and control extensive new findings can be anticipated in: 1. ferroelectric materials for memories, 2. materials the dielectric properties of which are highly sensitive to changes in ambients, and 3. an increased interest in the exploitation of mechanical-dielectric effects. A mention of ferroelectrics should not suggest that all that is needed are a few quick changes of atoms in the various lattice sites of the perovskite structure, followed by a functional life test program.

Ferroelectrics, unlike ferromagnetics, are per se 2-terminal devices and their properties are far more limited than are those of ferromagnetics. One need only consider the wide variations in magnetic susceptibility, coercivity, and magnetic saturation values and contrast these with the dielectric analog to see this. In fact, even though the ferroelectrics which are being marketed are used mostly in capacitors, the high temperature coefficient, their low electric strength, and low stability on aging suggest that many improvements are required at the materials level.

Many of the uses of such materials are planned for outer-space exploration, but few dielectricians may become involved in developing space ships and rockets. On the other hand, the manufacture of better products and more uniform items will come from developing instruments of control more suited and more apt for the job. The factory problems of today, while apparently mundane in comparison with those faced by the space conquerors, have the same three problems: detection, memory, and control. The judicious adaptation of dielectric effects to quality control offers a fruitful field for exploitation, for it is one of the few where a physical phenomenon has been related so aptly to the chemistry and uniformity of matter.

Let it not be thought that capacitance arrays should be installed on every factory operation. There are probably some being used today which might be removed with impunity. It would seem, however, that if the time-honored principles of dielectrics were re-evaluated in terms of their aptness and unique adaptability to automatic detection and control, improvements in product uniformity would well repay the effort.

References

1. Dielectric Behavior and Structure (book), C. P. Smyth, McGraw-Hill Book Company, Inc., New York, N.Y., 1955.

2. Theory of Electric Polarization (book), C. J. F. Bottcher. Elsevier Publishing Company, New York, N. Y., 1952.

3. Theory of Dielectrics (book), H. Frohlich. Clarendon Press, Oxford, England, 1949.

4. Dielectric Breakdown of Solids (book), S. Whitehead. Clarendon Press, 1951.

5. Dielectric Materials and Applications (book), A. von Hippel. John Wiley & Sons, Inc., New York, N. Y., 1954.

6. Electrical Insulation for Shipboard (book), G. L. Moses. John Wiley & Sons, Inc., 1952.

7. The Insulation of Electrical Equipment (book), W. Jackson. Chapman and Hall, Ltd., London, England, 1954.

8. Dielectrics, W. F. Brown. "Handbuch der Physik" (book), vol. 17, Springer-Verlag, Berlin, Germany, 1956.

9. Dipole Moments, L. E. Sutton. "Determination of Organic Structures by Physical Methods" (book), Academic Press, New York, N. Y., 1955.

10. Dielectric Absorption, M. Davies. *Quarterly Review*, Chemical Society, London, England, vol. 8, 1954, pp. 250–78.

11. K. W. Wagner. *Alta Frequenza*, Milan, Italy, vol. 20, 1951, pp. 3, 68.

12. M. Mandell. *Bulletin*, Société Chimique, Paris, France, July-Aug. 1955, pp. 7–8, 1018–25.

13. K. A. MacFadyen. *British Journal Applied Physics*, London, England, vol. 6, no. 1, Jan. 1955, pp. 1–7.

14. M. Bouix. *Annales Télécommunications*, Paris, France, vol. 7, pt. I, pp. 466–79; pt. II, pp. 497–512.

15. B. V. Hamon. *Circular*, Division of Electrotechnology, National Standards Laboratory, Sydney, Australia, Jan. 22, 1953.

16. C. P. Smyth. *Record of Chemical Progress*, Detroit, Mich., vol. 2, pt. I, 1950.

17. *Proceedings*, Institution of Electrical Engineers vol. 100, pt. IIa, 1953, p. 3.

18. *Insulation*. The Lake Publishing Corporation, Lake Forest, Ill.

19. Digest of Literature on Dielectrics. National Research Council, Washington, D.C.

◆

Discussion

R. W. Atkinson (General Cable Corporation, Westfield, N. J.): The author shows curves where the abscissa is frequency and goes to the high-megacycle range, the ordinate being called dielectric strength. In the range of the curves, large reduction of the ordinate takes place even for very low-loss material. This large fall-off in the voltage to produce failure is quite surely not dielectric failure in the usual sense. It is rather likely that, because of the increasing loss at the high frequencies observed, the failure is simply a result of heat. There is, in that case, no more reason to speak of dielectric strength than to speak of the dielectric strength of a wire which is melted if a sufficient voltage is placed across the terminals.

If these materials were tested in thin sheets the thermal time constant is quite short and the failure voltage will not be greatly affected by a time period of one minute or less. In any event, tests at the high frequencies and very low volts per mil should not be called tests of dielectric strength or other than a measure of dielectric loss unless this latter has been thoroughly eliminated by adequate analysis, theoretical or experimental.

L. J. Frisco (The Johns Hopkins University, Baltimore, Md.): Mr. Atkinson's remarks concern the author's statements relative to the drastic effect of frequency on the electric strength of solid insulating materials. The breakdown curves which were discussed during the oral presentation of the paper were based on the work of the late J. J. Chapman and his co-workers at the Johns Hopkins University. Data of this type have appeared in several publications.[1–3] Typical data for Teflon (PF), a low-loss material, and asbestos-filled diallyl phthalate (MDAP), a lossy material, are given in Table I.

Mr. Atkinson's remarks certainly apply to a material such as MDAP, which exhibits extremely poor behavior in the r-f range. It will be noted that in addition to a drastic decrease in electric strength, MDAP exhibits changes in the character of failure as the frequency of the applied voltage is increased.

In the case of the low-loss materials, such as PF, the observed decrease in electric strength is not accompanied by a change in the character of failure at the higher frequencies. With these materials, a sharp, pin-point type of puncture is obtained at all frequencies in the range of 60 cps to 100 mc (megacycles) when recessed electrodes are properly applied.[3] Furthermore, only minor variations in electric strength are observed when the breakdown tests are conducted over the temperature range of −55 C to 125 C. Attempts to explain these results on a simple thermal basis have been unsuccessful, a fact which Mr. Atkinson emphasized in his discussion of reference 1.

Other factors which depend upon the total number of cycles of applied voltage over a given test period have also failed to explain the variation of electric strength with increasing frequency. Consequently, the effect must be attributed to time-dependent mechanisms which are associated

Table I. Electric Strength, RMS Volts Per Mil

Frequency	PF	MDAP
60 cps.	850	836
1 kc.	808	465
38 kc.	540	145*
180 kc.	500	82†
2 mc.	375	21†
18 mc.	210	14†
100 mc.	143	*U*

* Puncture with attendant volume effect.
† Volume failure due to excessive heating.
U = Unstable condition due to excessive heating at very low stress.

with the breakdown process, and in those cases where the character of the breakdown is unaffected by increasing frequency, the measured values of stress must necessarily represent the electric strength of the material. A more complete discussion of the effect of frequency on the electric strength of solids may be found in the text and discussion of reference 1.

REFERENCES

1. DIELECTRIC FAILURE OF VOLUME AND SURFACE TYPES, J. J. Chapman, L. J. Frisco, J. S. Smith. *AIEE Transactions*, vol. 74, pt. I, July 1955, pp. 349–56.

2. DIELECTRIC BREAKDOWN TECHNIQUES, J. J. Chapman, L. J. Frisco, J. S. Smith. *The Electrochemical Journal*, New York, N. Y., vol. 102, Feb 1955, pp. 67–72.

3. DIELECTRIC STRENGTH OF SOLID INSULATION, J. J. Chapman, L. J. Frisco. *Electrical Manufacturing*, New York, N. Y., vol. 53, May 1954, pp. 136–43.

Subdivision of Telephone Traffic

IMRE MOLNAR
MEMBER AIEE

TRAFFIC DATA for engineering of telephone switching equipment is usually derived from traffic measurements extended over as large a number of circuit groups and as long periods of observation as practical. The quantities of equipment required in circuit subgroups are then based on the average traffic, prorated to the particular subgroup. The factors governing this averaging process will be analyzed under case I. The converse process, where the total traffic is estimated from observation on a few sample subgroups, is less often used in engineering equipment quantities. In latter instance, Case II, standard methods of statistical estimation are applicable, with additional adjustments which correspond to those discussed herein.

Case I

General Comments

It is important to realize that as telephone traffic is a random phenomenon with great variations in time and among individual sources, data derived from observations during short periods (say individual hours), and from small groups of sources, are useless in engineering. Instead, traffic measurements must be extended over a prolonged period, say many hours, and on many circuits, say over all lines of a central office, or at least over some of its major divisions. Similar consideration applies to the grade of service. For instance, in switching systems where attempts made during congestion are lost, the total attempts lost during prolonged periods of observation should be proportioned to the attempts made or carried during the same period. Whether the grade of service is to be judged for the over-all system or for each subgroup is a matter of definition, and will be discussed later.

However, even where data are obtained over a prolonged period on the over-all system, data of individual subgroups will not agree with an estimate derived by prorating the total system. For instance, the average busy-hour traffic may be originated or received by 10,000 telephone lines; that is, the total traffic measured many days during the hour of highest traffic each of those days, divided by the number of days of measurement. If the 10,000 lines are divided into 100 groups of 100 lines each, and the traffic in each of the 100 groups is measured during the same days but during the hour of heaviest traffic in each group, the average of each group will differ from the average of the first method divided by 100. Furthermore, if the averages of each group are added, the total will be higher than the average of the first method. However, even if the traffic in each group is measured during the hours when the combined traffic of the 10,000 lines is highest, the individual group averages will fluctuate considerably about this prorated over-all average.

There are two reasons for these phenomena. Even though the fact that there are pronounced busy hours in telephone traffic indicates that telephone calls are not placed completely at random, nevertheless the placement of calls by individual sources during certain periods can be considered as practically at random. Each group of sources will develop a busy hour of its own independently from other groups, and only dependent on the prevalent trend of its sources superposed on such random calling. The location of the group busy hours of the several groups will be pseudorandom relative to each other, though with a pronounced trend toward the location of the common office busy hour. This does not preclude the fact that the group busy hours are more pronounced than the office busy hour; on the contrary, these facts are interrelated.

To illustrate this point, consider Fig. 1 in which there are three groups each carrying the same busy-hour traffic, 4 erlangs. The first group has its busy hour between 9:30 and 10:30, the second between 9 and 10, and the third between 10 and 11. To simplify matters, assume that the traffic density (number of occupied devices) remains the same throughout the busy hour, i.e., 4 erlangs; and during the adjacent hours it drops abruptly by 25%, to 3 erlangs, even though, in actual serv-

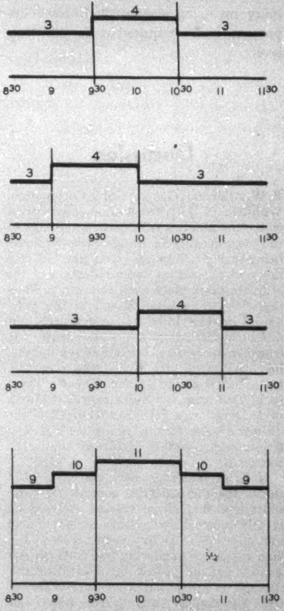


Fig. 1. Busy-hour traffic composition

Paper 57-809 recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted April 19, 1957; made available for printing October 8, 1958.

IMRE MOLNAR is with the General Telephone Laboratories, Inc., Northlake, Ill.

Vol.
Pt. 1
1959

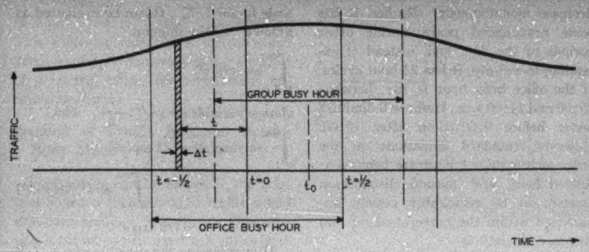


Fig. 2. Traffic density of a certain group in time

ice, the traffic density can be expected to decline gradually on both sides. If the three groups are combined, the over-all busy hour will be between 9:30 and 10:30 with 11 erlangs of traffic, while the sum of the three individual busy hours would be 12 erlangs, or 9% more. The two hours adjacent to the busy hour carry 9.5 erlangs each. The total combined traffic is 30 erlangs, which is, of course, the sum of the traffic in the three groups.

Now, two facts are immediately apparent. First, if the traffic of the combined busy hour is measured and an attempt made to infer from this the traffic of the individual groups, a value of 11/3 =3.67 erlangs is obtained instead of the true value, 4 erlangs; that is, a value which is 9% low. Thus, if one's engineering is based on the prorated average of the combined busy hour, one is caught short during the individual-group busy hour. Second, in the hours adjacent to the over-all busy hour, the combined traffic decreases only about 14%, while the individual-group traffic decreases 25%; therefore, the group busy hour is more pronounced than the over-all busy hour.

In the foregoing example it was assumed that the traffic in each group has the same magnitude, the only difference being the displacement of the busy hours relative to each other. Actually, the individual groups will carry differing amounts of busy-hour traffic, because it is practically impossible to assign to each group traffic sources which would generate exactly the same amount of traffic. Thus even if the busy hours should exactly coincide, the traffic of the individual groups will fluctuate about the prorated average of the combined group. The amount of fluctuation depends on the traffic-generating characteristics of the sources to the several groups with the objective of attaining as close a traffic balance as possible.

Busy-Hour Phase Shift

The phase shift between the busy hours of the subgroups is the more important of the two causes of traffic unbalance. In engineering switching equipment, allowance may be made for this, based on following reasoning.

Consider a large group or a whole central office, consisting of many subgroups each generating the same busy-hour traffic y. The office-busy-hour center point should represent the time $t=0$, while the subgroup-busy-hour center points t_0 are distributed according to some probability function $p(t)$. The traffic density in each subgroup is likewise distributed according to a function $f(t)$, which is assumed to have its maximum value at $t=t_0$, and to be monotonously decreasing and symmetric on both sides of t_0. Furthermore, $f(t)$ is assumed to be the same for all subgroups, except the shift of its crest to t_0. The integral of $f(t)$ between the limits $t_0-1/2$ and $t_0+1/2$ is y, the subgroup busy-hour traffic; while the integral between the limits $t_0+1/2$ and $t_0+3/2$, as well as between $t_0-3/2$ and $t_0-1/2$, is y', the subgroup adjacent-to-the-busy-hour traffic.

In Fig. 2 the function $f(t)$ of a certain group is shown, centered around its particular t_0. During a brief interval Δt at an arbitrary time t in the office busy hour, this group will generate traffic $f(t-t_0)t$; therefore the total traffic of the group during the office busy hour will be

$$\int_{-1/2}^{+1/2} f(t-t_0)dt$$

Assuming again that the function $f(t)$ is the same for all groups, the value of the group traffic during the office busy hour will depend on the location of t_0. As has been shown, t_0 varies according to a probability function $p(t_0)$, so the traffic expected in an arbitrary *group* during the *office* busy hour will be

$$\int p(t_0)dt_0 \int_{-1/2}^{+1/2} f(t-t_0)dt \qquad (1)$$

The limits of the first integral depend on how far the group busy hours stray from the office busy hour. They will, of course, in any case be contained within a 24-hour period. Ordinarily, for traffic groupings of practical interest, almost all group busy hours remain within less than 2 hours on either side of $t=0$ and therefore little error will be incurred by extending the limits of integration to infinity.

To solve equation 1 the traffic distribution $f(t)$ of a subgroup should be known. According to statistical evidence, the decline of the group busy-hour traffic to adjacent hours is gradual. If the group busy-hour traffic is y, and the traffic during the hour on each side is y', the decline is about 35% for 1-erlang traffic, 15% for 10 erlangs, and 12% for 100 erlangs. To simplify evaluation of equation 1, assume that during the group busy hour $f(t)$ has a constant value y, and it then drops abruptly to y'. There is also evidence that the decline of traffic during the second hour away from the busy hour is even less, and as its influence on the office busy-hour traffic is small anyhow, assume that everywhere outside the

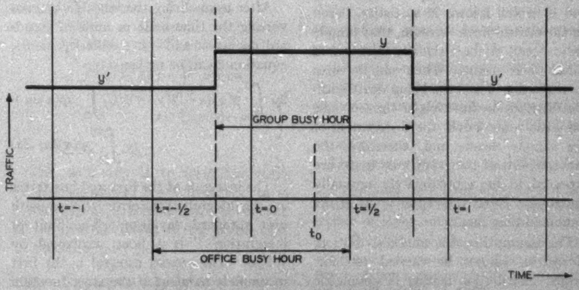


Fig. 3. Relation of office to group busy hour

group busy hour $f(t)$ remains constant, y'.

From these considerations and from Fig. 3, estimate the contribution of a group to the office busy-hour traffic as $y(1-t_0)+y't_0$ when $-1<t_0<+1$, and y' when t_0 lies beyond. The traffic expected in the group during the office busy hour will therefore be

$$\int_0^1 p(t)2y(1-t)dt+\int_0^1 p(t)2y'tdt+\int_1^{\mathrm{inf}} p(t)2y'dt \quad (2)$$

To solve equation 2 one must also know the probability distribution of the group-busy-hour location $p(t)$. From statistical evidence it appears that in local telephone traffic the group-busy-hour center points are approximately normally distributed about the office-busy-hour center point, with a standard deviation of s hour. Therefore, one hour is $1/s$ standard deviations.

There is also some theoretical justification for the normal distribution of the group-busy-hour center points. Consider again a central office consisting of 100 groups of 100 lines each. Each of the 10,000 lines is used at random, or at least according to some frequency function about which little is known. These 10,000 lines form a universe, from which 100 samples are picked in sets of 100 lines each. (Although irrelevant for this illustration, it is assumed that each 100-line sample originates the same traffic volume.) It may likewise be supposed that there is a "busy hour" for every line, also according to some frequency distribution over the universe, of which, also, little is known. The center points of the busy hour of each of the sample groups are the mean values created by the individual-line busy hours within the group, and the center point of the office busy hour is a mean value created by the distribution of the center points of the group busy hours. These group busy hours are distributed according to some frequency function, and it is well known in statistics, based on the central limit theorem, that sample means tend to be normally distributed about their mean. When sample sizes on the order of 100 are being dealt with, the distribution function of the universe has little control over the distribution of the sample means and, therefore, the center points of the group busy hours are expected to be approximately normally distributed about the center point of the combined busy hour.

The assumption of normal distribution, of course, cannot be carried too far. Traffic distribution neither is symmetric about the office busy hour, nor does it decrease monotonously. Rather, it has some pronounced peaks during other periods of the day, and, instead of extending to infinity, it has 24-hour cycles. If the office busy hour is, say, between 10:00 and 11:00 a.m., traffic will diminish faster before 9:00 than after 12:00. However, standard deviations on the order of less than 1 hour are being considered here, and normal distribution appears to be reasonably closely approached within the range of about two standard deviations.

The value of the standard deviation s is obviously a function of the traffic: If the groups handle large traffic volumes, their busy hours will be clustered more closely together than would be the busy hours of groups with less traffic. Consider a population consisting of the center points t_0 of the group busy hours. According to the previous assumptions, each group carries the same traffic y, and t_0 has a Gaussian distribution with a mean value 0 and variance s^2, where $s=s(y)$. A large number of sample groups, say 100, are now chosen from the population. The sample means can be expected to be near 0, the sample standard deviation close to s, and the sample distribution to be normal. A second sample of 100 groups is now taken, and by random addition of the two samples a new sample of 100 groups is obtained, each carrying $2y$ erlangs of traffic. The mean of the new sample will be near 0, the standard deviation near $s/\sqrt{2}$, and again the distribution will be normal. Similarly, combining n sets of sample groups will yield a new set of sample groups carrying a traffic of ny erlangs, with mean value 0 and standard deviation $s(ny)=s(y)/\sqrt{n}$. The only assumptions made were: 1. a normally distributed population of a very large number of groups each carrying the same traffic, and 2. a very large number of sample groups, yielding a large number of samples even after n groups were combined.

After normalizing the variable by converting the time scale in units of standard deviation and other rearrangements, equation 2 can be replaced by

$$2y\int_0^T p(u)du-[2(y-y')/T]\int_0^T up(u)du+2y'\int_T^{\mathrm{inf}} p(u)du \quad (3)$$

The integrals of the first and last terms are the integrals of the error function, with unit standard deviation. The limit of integration T is 1 hour, expressed by $1/s(y)$. The second integral is the first incomplete moment of the error function from 0 to T standard deviations, with a scale factor $1/T$. It can be evaluated as follows. In the integral

$$\int_x^{\mathrm{inf}} vdv/\exp(v^2)$$

change variable $v=x\sqrt{(1+r)}$. Thus

$$\int_x^{\mathrm{inf}} vdv/\exp(v^2)=1/2\exp(x^2)\int_0^{\mathrm{inf}} d(rx^2)/\exp(rx^2)=1/2\exp(x^2)$$

After replacing the variable v by $u/\sqrt{2}$ and multiplying both sides by $1/\sqrt{(2\pi)}$, the right side becomes the error function with unit standard deviation $p(u)$, thus

$$\int_0^T up(u)du=p(0)-p(T) \quad (4)$$

Designating

$$2\left(\int_0^T p(u)du-(1/T)\int_0^T up(u)du\right)=2\left(\int_0^T p(u)du-(1/T)[p(0)-p(T)]\right)$$

by A, equation 3 becomes

$$Ay+(1-A)y'$$

The last expression represents the contribution of a subgroup to the office busy-hour traffic, when the subgroup busy-hour traffic is y and the subgroup traffic adjacent to the busy hour is y'. If the combined traffic of all n subgroups during the office busy hour is Y, the prorated amount applicable to one subgroup is

$$Y/n=Ay+(1-A)y' \quad (5)$$

In the usual case Y is known from estimate or measurement, and y can be calculated by increasing Y/n by the percentage

$$a=100(y-Y/n)/(Y/n) \quad (6)$$

If, however, the subgroup busy-hour traffic y is known and its contribution to the combined busy hour is to be estimated, y should be decreased by the percentage

$$a_1=100(y-Y/n)/y \quad (7)$$

Then a and a_1 can be tabulated against Y/n and y, respectively, or instead, corresponding values of Y/n and y can be shown. Above about 2 or 3 erlangs an interchange of a and a_1 is within engineering accuracy.

Obviously, the closer y' approaches y (which actually it does when traffic increases), the smaller adjustment a is required. In the trivial case $y'=y$, equation 6 gives $Y/n=y$, as it should, because when the traffic volume embraces a large proportion of the total traffic

(n approaches 1), Y also approaches y. This conforms with the experience that when a group embraces a major division of a central office, the adjustment becomes negligible.

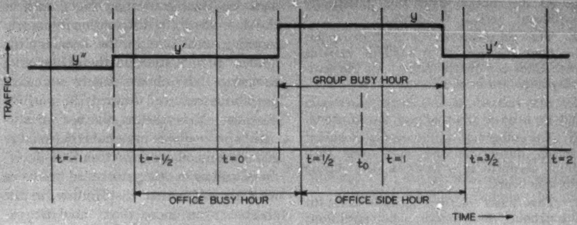


Fig. 4. Relation between busy hour and side hours

Decline of Traffic Adjacent to the Busy Hour

Equation 5 can be solved if the relation between the busy-hour traffic y and side-hour traffic y' is known. This relation is not inherently fixed, but varies from case to case, and, in particular, it will depend on the amount of traffic y, as already implied. The decline becomes less pronounced with increasing traffic. For any specific application, the relation between y and y' can be established reliably only by a traffic study. However, a general idea can be obtained by considerations similar to those in the preceding section.

Assume again that the traffic in the side hours of a large group is generated by its subgroups. The contributions of the subgroups are again shifted in phase, just as were their contributions to the over-all busy hour. And, although the exact amount really is not very important, assume that subgroup traffic 2 hours removed from the subgroup busy hour declines from y' to $y''=y'(1+y'/y)/2$.

Fig. 4 illustrates this case, except that for simplification the second decline just referred to will be considered only when t_0 is less than 0. When t_0 is above $t=2$, its effect is insignificant. The traffic each group contributes in the hour adjacent to the over-all busy hour is

y'' if $t_0<-1$

$y'(t_0+1)-y''t_0$ if $-1<t_0<0$

$yt_0+y'(1-t_0)$ if $0<t_0<1$

$y(2-t_0)+y'(t_0-1)$ if $1<t_0<2$

y' if $2<t_0$

where t_0 is again the center point of the group busy hour and $t=0$ is the center point of the over-all busy hour. The expected contribution of a group to the traffic of the side hour is, therefore,

$$y''\int_{-\inf}^{-1}p(t)dt+(y'-y'')\int_{-1}^{0}tp(t)dt+ y'\int_{-1}^{0}p(t)dt+(y-y')\int_{0}^{1}tp(t)dt+ y'\int_{0}^{1}p(t)dt+(2y-y')\int_{1}^{2}p(t)dt- (y-y')\int_{1}^{2}tp(t)dt+y'\int_{2}^{\inf}p(t)dt$$

$$=y'(1+y'/y)/2\left(1/2-\int_0^1 p(t)dt+\int_0^1 tp(t)dt\right)+(y-2y')\int_0^1 tp(t)dt+ 2y'\int_0^1 p(t)dt+(2y-y')\int_1^2 p(t)dt- (y-y')\int_1^2 tp(t)dt+y'\int_2^{\inf}p(t)dt \quad (8)$$

The frequency function $p(t)$ will be that used in the preceding section, with the same variance $s(y)^2$. Therefore, after normalizing equation 8 by a change of variable, the following is obtained:

$$y\Bigg(2\int_T^{2T}p(u)du+\Bigg[(1/T)\Bigg(\int_0^T up(u)du-\int_T^{2T}up(u)du\Bigg)\Bigg]+y'\Bigg[2\int_0^T p(u)du-\int_T^{2T}p(u)du+\int_{2T}^{\inf}p(u)du+(1/T)\times\Bigg(\int_T^{2T}up(u)du-2\int_0^T up(u)du\Bigg]+(1+y'/y)/2\Bigg[1/2-\int_0^T p(u)du+(1/T)\int_0^T up(u)du\Bigg] \quad (9)$$

After a few rearrangements, equation 9 can be expressed as

$$Y'/n=Cy+Dy' \quad (10)$$

where

$$C=2\int_0^{2T}p(u)du-2\int_0^T p(u)du+(1/T)[p(0)-2p(T)+p(2T)]$$

$$D=1/2+\int_0^T p(u)du-2\int_0^{2T}p(u)du-(1/T)[2p(0)+p(2T)-3p(T)]+(1+y'/y)/2\left\{1/2-\int_0^T p(u)du+(1/T)[p(0)-p(T)]\right\}$$

$$=1-C-E$$

The next step is to form the ratio $(Y/n)/(Y'/n)=Y/Y'=k$ from equations 5 and 10, as a measure of the decline of the over-all traffic in the hours next to the office busy hour

$$Y/Y'=\Big(Ay+\Big(1-A\Big)y'\Big)\Big/(Cy+(1-C-E)y')=k$$

where

$$E=(1+y'/y)/2\left\{1/2-\int_0^T p(u)du+(1/T)[p(0)-p(T)]\right\}$$

and

$$y'=y/(1+[k(1-E)-1]/(A-kC) \quad (11)$$

The ratio y'/y needed for E can be estimated by successive approximation by first assuming E to be zero and computing y' for a few selected values of k. The next step usually provides a sufficiently close value of E. As the parameters A and C and E are tabulated functions of the normal distribution and depend on y only through $T=1/s(y)$, they can readily be computed before y' is known. Parameter k may be considered a constant of a given universe of line groups, say for a certain central office, when it is the ratio of the office busy-hour traffic to its adjacent-hour traffic.

Review of the Conditions

Before proceeding with the numerical evaluation of the results, the basic assumptions will be reviewed. It may also be of interest to note that the preceding discussions did not employ the usual paraphernalia of telephone traffic problems: statistical equilibrium, holding-time distributions, Poisson process, congestion theory, etc. Instead, the following few assumptions were made as to the gross statistical distribution of the traffic over a period of several hours, and then the results were obtained directly by the elementary analytic methods of sampling, using only the extensively tabulated error function.

The assumptions were as follows:

1. The traffic of a large unit, for instance of a large central office, is the sum of the traffics of many subgroups of the unit.
2. The traffic of the unit as well as the subgroups, integrated over 1-hour periods, has a definite maximum in a prolonged period, say within 24 hours.
3. The traffic of every subgroup, as well

as its distribution over the statistically significant period, is identical.

4. The subgroup-busy-hour center-point locations conform to a random variable with Gaussian distribution about the unit-busy-hour center point.

5. The variance of this distribution is in inverse ratio to the traffic of the subgroup.

6. The subgroup traffic has a constant density during its busy hour, and then drops abruptly to a lesser density during the side hours.

7. This lesser density remains constant throughout the statistically significant range of the side hours.

8. The population of subgroups contains a sufficiently large number of subgroups so that the results will be asymptotically exact.

Granting the foregoing assumptions, there remains only to specify two parameters for the numerical evaluation: the variance of the distribution, and the ratio of the traffic in the busy hour to the traffic in the side hours. Neither of these two has inherently universal properties, nor are there any probability considerations to be of help. They will vary between individual applications; from a traffic study for each case, they should preferably be established by statistical methods. Traffic characteristics of a central office serving predominantly business establishments differ from those of a central office serving only residences. Traffic characteristics of a college town differ from those of a seaport. Paystation groups have their own peculiar traffic distribution. Toll traffic has different characteristics from local traffic. Individual toll-circuit groups have pronouncedly separated busy hours, particularly where the terminals are in different time zones. With such diverse conditions it would seem impossible to construct universally applicable adjustment factors for subgroupings. However, there are enough applications of generally similar characteristics to permit the computation of an average set of adjustment factors, sufficiently close for engineering use where proper statistical data are not available. The usual adjustment rarely exceeds 10% to 15%, and if adjustments are applied conservatively, second-order accuracy is not warranted. It is more important to be conscious of the need of an adjustment and to apply the approximate amount, than to seek dubious precision.

Assumption 1 is self-evident, and assumption 2 is an established fact. Assumption 3 is a temporary expedient which was used to appraise the contribution of the busy-hour phase shift, and will be dropped in later sections. Theoretic justification of assumption 4 has already been attempted, in spite of some undeniable contrary facts, such as the skewness of the distribution from early morning until late in the evening, the 24-hour periodicity, and the existence of subgroup busy hours many standard deviations removed toward the positive abscissa. Nevertheless, the test for normality on available material is reasonably good within about two standard deviations, and as in other statistical processes with a quasi-normal distribution, in the absence of a more exact analytic expression, the error function yields satisfactory results for engineering purposes. Actually, the subgroup busy hours will usually be contained within about a 4-hour period during the morning, and the number of those which fall beyond are fewer than what should be contained in the tails of the normal distribution. Therefore, the theoretic adjustment factor a is somewhat overestimated, which will tend to compensate for the errors due to assumptions 6 and 7.

Theoretical justification has also been attempted for assumption 5, but its statistical verification is rather difficult because it requires substantial data covering a wide range of subgroup traffic values, and the available measuring technique is complex and tedious. The best that can be said is that the scant material available to the author does not refute the assumption. This uncertainty is unfortunate, and because of the two parameters the standard deviation is the more decisive on the shape of the functional relation between traffic and adjustment a. When the standard deviation is small, the integration over the office busy hour will include practically all subgroup busy hours, and equation 5 factor A will approach unity rapidly with increasing subgroup traffic y on two accounts: more rapid convergence of the error integral, and the scale factor $1/T$. Side-hour traffic y' will have a diminishing importance, and the need for adjustment a will disappear at relatively low subgroup traffic volumes. Conversely, when the standard deviation is large, the side-hour traffic will exert more influence, and larger adjustments a are required.

It has been seen that if the standard deviation $s(y_1)$ for a certain traffic y_1 is known, then the standard deviation for traffic value y is $s(y)=s(y_1)/\sqrt{(y/y_1)}$. If one knows the value $s(y_1)\sqrt{y_1}=1/r$ from a certain set of observations, $s(y)$ can be determined for any value of y from assumption 5 as $s(y)=1/r\sqrt{y}$, provided that y and y_1 belong to sets of samples of the same normally distributed population, and that $s(y_1)$ is a satisfactory estimate for the standard deviation of the population. Then parameter T of the error function is

$$T=1/s(y)=r\sqrt{y} \tag{12}$$

The value r should preferably be determined from a study of the distribution of the center points of a large number of group busy hours in some typical applications with relatively small values of traffic, such as on 100 or more groups of 100 lines each within a central office. Even though the values of r obtained this way may not have universal validity, the average from a sufficient number of studies will permit the construction of a curve $a=f(Y/n)$, which if judiciously applied should be satisfactory.

From studies the author has made on the terminating-traffic distribution of large central offices, as well as from material available elsewhere in the literature, it appears that for subgroup traffic on the order of 2 to 2.5 erlangs the standard deviation may vary between 0.55 and 0.65 hour, with a range for r between 1.2 and 1.0 if the traffic is expressed in erlangs. Conservatively, 1.0 was chosen for the computations in the present study. For different traffic characteristics, for instance in toll traffic where local time differences increase the standard deviation, the adjustment factors a which are suggested here may need to be slightly increased.

Assumption 6 is obviously contrary to the facts, since the traffic density shows a marked decline in both directions already during the busy hour, with a gradual transition toward the side hours. Nevertheless, the assumption is tenable, for any better interpolation would unduly complicate the computations without sufficiently increasing the accuracy of our result. Assumption 6 slightly overemphasizes the contributions of the group busy hours near the office busy hour, and these hours carry more weight due to the characteristic of the error function $p(u)$. Therefore, the value of a is slightly underestimated but the error is small compared with the uncertainty already present in the choice of r. The error is greater when the difference between y and y' is greater, that is, for small values of y. By the choice of a low value for r and, as will be shown later, by the character of the ratio y/y', such errors are already sufficiently compensated.

Assumption 7 is likewise contrary to the facts, since the traffic density declines steadily during the side hours, at least within a range of several hours. Here, again, the assumption simplifies computations without markedly debasing the accuracy. The error is due to extension of the integration over the last term in equation 3 to infinity, instead of only to $2T$, plus successive terms with steadily diminishing traffic values. Now, for traffic values above about 1 erlang the difference between the integral to infinity and to $2T$ is less than 2%. But since this integral even at 1 erlang only represents about one third of the value of equation 3, the use of y' over the tails of the distribution instead of a, say, 20% to 30% lesser value, would only reduce the value of equation 3 by a fraction of 1% at 1 erlang, and would be completely insignificant at higher traffic values. In general this error will likewise tend to underestimate the adjustment factor a.

Assumption 6 would also tend to give a somewhat high value for y', due to equation 8, which again is insignificant when the traffic is not too small. Small values $(y-y')/y$ tend to reduce adjustment a, but since it has only a second-order effect, and even for low traffic values the error is small anyhow, a conservative choice of the parameters will compensate adequately. The errors incurred by assumption 7 in the computation of y' were for all practical purposes compensated by the term E.

Assumption 8 has not been explicitly stated in the derivation of the equations; however, its oversight may lead to paradoxical results. For instance, consider a central office consisting of 100 subgroups of 100 lines each, with a total busy-hour traffic of 250 erlangs. The prorated traffic per subgroup is 2.5 erlangs, and if say, a 12% adjustment is applied, the subgroup busy hour will be 2.8 erlangs. If the theory is correct, traffic measurement should confirm it. Now, instead, take the same 10,000 lines and divide the office into ten groups of 1,000 lines each, with a prorated traffic of 25 erlangs per group. The correct adjustment for such a group is 2%, and the group busy-hour traffic will be 25.5 erlangs. If in the next step only one of these 1,000-line groups is considered out of context with the central office of which it is only a part, and it is divided into ten subgroups of 100 lines each, the prorated traffic per subgroup is 2.55 erlangs. Applying the same approximate correcting factor used previously (12%), the subgroup busy-hour traffic should be now 2.86 erlangs instead of the previous 2.8 erlangs. Obviously the subgroups remained the same, and thus a dilemma arises.

The answer is in assumption 8. In integrating the distribution function as well as in deriving equation 12, an infinite population of normally distributed subgroups was assumed, and the expression obtained for adjustment a is an asymptotic one with the proviso that in the subdivision Y/n, n is a very large quantity, which means that a large volume of traffic was subdivided into many components. When the central office was subdivided into 100 subgroups, 2.8 erlangs were obtained in each, in accordance with the hypothesis. However, when the central office is divided into ten groups, these will not carry equal amounts of traffic in their group busy hours, but will have a distribution similar to a set of ten random samples of ten random items each from the population of the subgroups. Now it is known from an exact sampling theory that in the distribution of samples containing only a few items the expected standard deviation of the samples is less than the standard deviation of the population. For instance, in samples containing only ten items of a Gaussian population the expected standard deviation is about 7% less than the standard deviation of the population. It has been shown that the adjustment a required for subgrouping diminishes with decreasing standard deviation. There are exact sampling theories available in statistics where the sample sizes are taken into consideration for the estimation of the parameters in a distribution. Indeed, it can be shown by more detailed analysis that adjustment a becomes less when the overall traffic is divided into fewer components; that is, the smaller n is in Y/n. The construction of charts with n as parameter, as well as their use, becomes unduly involved, and the accuracy of the asymptotic approach suffices for all practical purposes.

However, in engineering subdivided traffic, it is advisable to consider whether the traffic which is to be divided embraces a whole population, or is only a sample of the population and whether as such has been derived already by a process of subdivision. In the latter case, and referring again to the previous example, the 25.5 erlangs, which were assumed to be the group busy-hour average traffic, should have been reduced to the office busy-hour traffic of 25 erlangs before further division into ten subgroups each.

Computation of the Correcting Factor

The last step is to assign a value to the parameter k in equation 11. As will be recalled, if k is known, the traffic y' in the hours next to the subgroup busy hour can be computed, based on assumption 8. The constant k is the ratio of the office busy-hour traffic Y to its adjacent-hour traffic Y', and, of course, will vary in different applications. The problem again is to choose a value which would fit most of the important cases. Obviously the higher the value of k, the greater the spread between the busy hour and the side hours, and the greater adjustment a is necessary. Fortunately, the office busy-hour traffic values Y and Y' are comparatively simple to determine from the traffic volumes of large central offices, and $k=1.12$ appears to be a suitable value. The per-cent decline of traffic $b=100(y-Y')/y$ from the busy hour to the side hours is plotted with $k=1.12$ in Fig. 5.

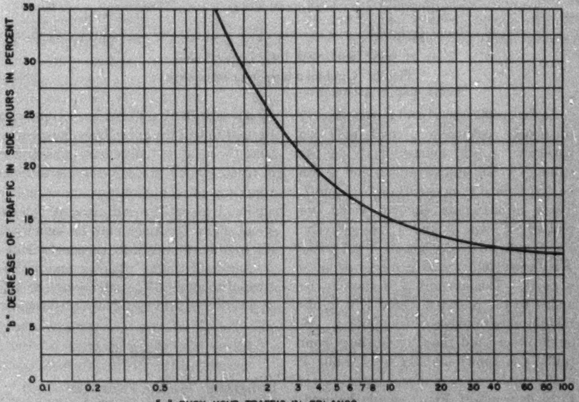


Fig. 5. Per-cent decline of traffic from busy hour to side hours

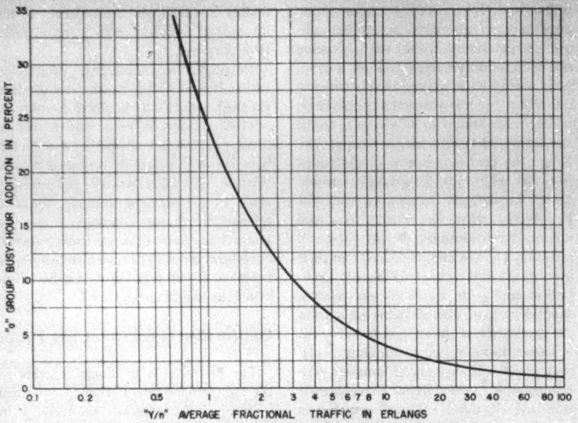


Fig. 6. "Y/n" subgroup busy-hour traffic adjustment

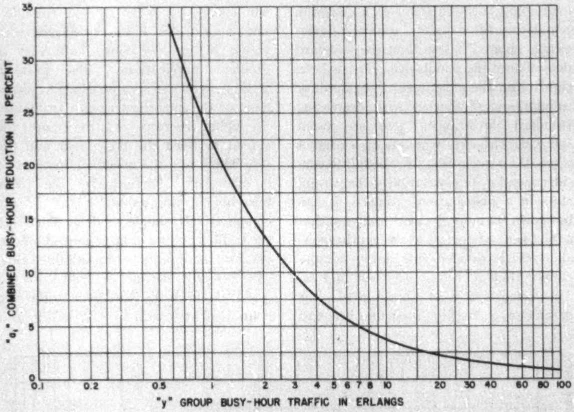


Fig. 7. Combined busy-hour adjustment

As the coefficients A and C were already previously calculable, based only on parameter r and independently of the value of k, the adjustments a and a_1 can be determined from equations 5, 6, and 7. These adjustments (in per cent) are plotted in Figs. 6 and 7, respectively. Fig. 6 is to be used when the total traffic Y is known, and it is to be subdivided into n parts. The amount a is then to be added to the prorated traffic Y/n to obtain the subgroup busy-hour traffic y. If, on the other hand, the subgroup busy-hour traffic y is known and the total traffic Y of the combination of n such subgroups is to be determined, Fig. 7 should be used. The adjustment a_1 is to be subtracted now from y and the remainder multiplied by n to obtain Y. These figures may also be used to find the group busy-hour traffic from the measured office busy-hour traffic of trunk groups or vice versa, for tandem trunking, alternate routing, and similar applications. The adjustment above about $Y/n=5$ erlangs may be somewhat low in some applications, such as for long-distance toll traffic, where it may be increased to about 6% at 10 erlangs, to 4% at 20 erlangs, and to 2% at 50 erlangs.

Case II

Subgroup Traffic Variation

The discussion of Case I explains the general busy hour in a telephone office as a synthesis of many individual subgroup busy hours shifted in phase relative to each other. With the adjustments specified in Figs. 6 and 7, the busy-hour traffic of a *subgroup* can be estimated if its *parent-group* busy-hour traffic is known, or conversely. This is very useful in many applications where such individual information is available.

In the majority of applications for local switching, however, only the over-all office busy-hour traffic Y may be available, and in this case the individual subgroup traffic can be estimated by Fig. 6 only if, according to assumption 3, all subgroups carry equal traffic. Now, such even subdivision of traffic is attainable only in special applications, such as in an automatic call-distribution system to a team of operators where the work-time of calls is approximately equal, but in such case not even a phase-shift adjustment is permissible. In more common applications, however, even subdivision is impossible. There are cases where it is inherently impossible, such as when a large amount of traffic is to be split into several unequal trunk groups. There are other cases where equalization has great practical difficulties, for instance when the bulk terminating traffic in an office is to be channeled to, say, 100 connector groups of 100 terminals each. In the administration of number assignments great care usually is taken to select the traffic characteristics of individual customers with the object of traffic equalization between the connector groups, and, to some degree, the success depends on the effort devoted to the selection. But no matter how perfect the administration is, there is a limit, because characteristics change in time, particularly when vacated terminals must immediately be reassigned, and also because number changes are to be avoided.

From what has been said it is apparent that in the first case, the varying sized trunk groups, each group must be individually appraised, preferably by individual traffic measurements. In the second case, the connector grouping, only an average estimate of the variation can be made, which may be too high or too low, depending on the effectiveness of the terminal assignment in the particular application.

If the subgroup traffics vary about their average, it may not be correct to apply the average adjustment associated with the average subgroup traffic. Fig. 7 shows that the adjustment decreases with increasing traffic, but if drawn on a linear scale, the rate of change is reasonably constant within a moderate range, and it

Vol.
Pt. 1
1959

likewise decreases with increasing traffic. This means that if the variation of subgroup traffic is moderate and symmetric about the average, it is satisfactory to work with the average adjustment. Actually the variation of traffic is such that the groups with lower-than-average traffic are clustered closer to the mean, while those with higher-than-average traffic stray farther away, which will tend to compensate for the decreasing slope of the adjustment.

These relations can also be considered through equation 5, where $Y/n=yA+y'(1-A)$. Of interest here is the constancy of $d(Y/n)/dy=(y-y')\,dA/dy+(1-A)dy'/dy+A$ within a reasonable range or that the second derivative $2dA/dy-(y'+dA/dy)(dy'/dy)+(y-y')(d^2A/dy^2)+(1-A)(d^2y'/dy^2)$ should be insignificantly small. These can be verified by inspection of the second differences and also by intuition: in $d(Y/n)/dy$ the first factor of the first term increases with increasing y, the second factor decreases probably to a somewhat lesser degree, both factors in the second term decrease, and the third term increases about the same degree as the second factor in the second term decreases. Thus it can be expected that the whole expression tends to be constant within a moderate range. Indeed, three rather extreme cases with 100 subgroups may be calculated: in the first, all 100 groups carry the same traffic $y=2$ erlangs; in the second, 50 groups carry 2 erlangs each, and two sets of 25 groups each carry 1 and 3 erlangs respectively; in the third, 60 groups carry 2 erlangs each, 30 groups 1.5 erlangs each, and 10 groups 3.5 erlangs each; and the results of the three will be indistinguishable. As a matter of fact, one more case can be calculated: 50 groups carry 1 erlang traffic each, and 50 others carry 3 erlangs each; the result will be only about 1/2% higher than the previous three, which is insignificant when one considers the accuracy with which the graphs can be read.

The Grade-of-Service Objectives

Apparently assumption 3 may be dispensed with since Figs. 5, 6, and 7 give the information necessary to estimate group busy-hour traffics, office busy-hour traffics, and side-hour traffics, subject, of course, to the validity of the other assumptions and the propriety of the empirical parameters r and k. Under these conditions it should be possible to obtain the expectation of the traffic by statistical methods.

However, traffic data are only the means to estimate the amount of equipment needed, and the equipment is only the means to provide service to the user. As in all branches of engineering, it is useless to attempt to specify equipment for a certain demand unless the service to be given is also considered. But it is interesting to note that while the demand (in this case the traffic) is something real and the equipment, of course, something very tangible, the service, on the other hand, is something elusive. It is a simple matter to express the grade service in concrete terms, and to construct a mathematic model or measuring means to verify whether this grade of service is satisfied or not by the equipment in any one situation. But one is confronted with great obscurity if he tries to find out by an objective measure what grade of service is the correct one for the specific type of service as far as the user of the service is concerned.

How is the grade of service in telephone traffic established? One may take a scientific approach, based on the fact that improved service creates an increasing volume of business, and continue to improve service until the point of diminishing return is reached when the cost of improvement is greater than the incremental revenue it creates. Or one may say that the service is to be given that the customer demands, but then he is on very uncertain grounds. The customer's reaction is very different depending whether he is in a situation where he has to put up with poor service regardless how much he complains, or where his complaints are heeded even though generally the service is considered satisfactory. To a large extent the grade of service is established by tradition based on the conditions prevailing in a particular service industry and to which the public became accustomed. A person may continue to patronize a transportation line, but would flood the Public Service Commission with complaints if a similar grade of service would be given by the telephone company. He may do business with a telephone order house or a ticket reservation office, but would not tolerate the same delay in answer by a toll operator. Many of the accepted grades of service were established by individual judgment, sometimes just intuitively or even emotionally, but once they entered the public domain they tend to perpetuate themselves—with a trend toward continual improvement.

In general, the grade of service provided by telephone systems is among the highest in any industry, and it appears to be sound, judging from their phenomenal growth and the satisfactory return on the investment at reasonable rates.

The purpose of the foregoing digression is to make it clear that there is nothing objective behind the apparent exactitude of the grade of service or the various congestion theories. Their purpose is to establish a common measure for the comparison of service results, and nothing more. The choice of one or another grade of service must be made on considerations unrelated to the mathematic problems of switching and congestion. The grade-of-service objective may be expressed in an almost unlimited variety of ways, of which the following are typical:

1. In a set of subgroups, the grade of service given to the subgroup with highest traffic during its average busy hour should not be worse than the specified amount.
2. In a set of subgroups, the average grade of service given to all subgroups in a switching rank during their respective busy hours should not be worse than the specified amount.
3. In a set of subgroups, the average grade of service given to all subgroups in a switching rank during a specified period, say over 24 hours of the average business day, should not be worse than the specified amount.
4. For a central office, the grade of service given by a switching rank to the combined traffic during the average office busy hour should not be worse than the specified amount.

Requirement 1 is obviously extravagant. With the usual variations among subgroups there are only very few whose traffic exceeds the average by, say, 30%, and to take care of these few may mean a general increase in cost of perhaps nearly 30%. Requirement 2 has a bearing on the subgrouping adjustment and will be discussed first in some detail, with the other specifications to be taken up afterward.

Behavior of the Loss

If objective 2 is accepted, one can expect that subgroups which carry more-than-average traffic during their subgroup busy hours will have a poorer than average, while those with lower-than-average traffic will have a better-than-average grade of service, but the grade of service for the totality of subgroups will be the specified amount. For the time being, consider only the grade of service in a single switching rank. As a measure for the grade of service take the amount of attempted traffic "lost" because of insufficient number of trunks, and calculate it according to Erlang's theory. The latter choice was made as a matter of convenience, since the result would not be noticeably different by any other of the accepted theories. For the present purpose, one may also disregard subtleties in definitions applied to Erlang's theory.

Assume then that an office with 100 groups carries traffic $y=220$ erlangs, so $Y/n=2.2$, and with $a=13\%$, the traffic per group is about 2.5 erlangs. Specify a grade of service $P=0.01$, and for $y=2.5$ at $P=0.009983$ Erlang's theory specifies $N=7$ trunks. Assume also that the maximum group busy-hour traffic observed in any of the groups is 3.7 erlangs, and the minimum is 1.7 erlangs. If seven trunks each are assigned to all groups, in the 3.7-erlang group they will give a grade of service $P=0.048270$, and in the 1.7-erlang group $P=0.001488$. These two average $P=0.024879$, that is, 2.5 times the objective. The reason for this is that loss equations are highly nonlinear. Loss increases steeply with increasing traffic, so the average loss will be considerably higher than the loss computed for the average traffic.

Based on specification 2, the groups with high traffic will suffer higher losses and those with small traffic, lower losses, but the objective here is to keep the average loss to 1%, which means that the sum of all individual-group losses should be 1% of the sum of all individual-group busy-hour traffics, or 2.5 erlangs. To accomplish this, the number of trunks must be increased to more than seven. To include the effect of group variation in the congestion equation is impractical. Instead, the average group busy-hour traffic figure can be increased enough to obtain from the standard tables the correct number of trunks yielding the specified loss. This proposed second adjustment c is in a way fictitious. A real increase in traffic over y does not take place, and therefore is not measurable. Its only purpose is to compute the number of trunks required to satisfy specification 2.

Variation in Subgroup Traffic

To obtain adjustment c, one must know the characteristics of the subgroup traffic variations. As already mentioned, there is nothing inherently fixed in these variations, but they change from situation to situation, dependent on the success in the traffic-equalization attempt. The traffic a large-city local central-office unit switched to 95 groups of stations, with an average of about 96 stations per group, was studied for a year. This particular central office was chosen because the character of the stations was well balanced between business lines, residential lines, individual lines, party lines, and pay-station lines, and the administration of load assignment was somewhat better than average, and also because during the period of the study there was less than ±0.4% change in the number of stations connected to the unit. While the administration may have been better than average, there is little choice in subsequent reassignments in an office over 10 years in service and filled to capacity. Furthermore, due to the reluctance for number changes, unbalance is quite pronounced in terminating traffic while in originating primary groups the degree of load balance depends on the effort devoted to line rearrangements. The average number of calls received per station in the group busy hours was 0.59, and the average group busy-hour traffic per group was 2.572 erlangs with 0.382-erlang standard deviation. The characteristics of larger traffic values were obtained readily by random combination of the subgroups. Other studies of lesser scope, and also evaluation of material found in the literature, confirmed that the conclusions derived from these studies can be generalized for the average application. There are some common characteristics in all these distributions. For instance, it appears that the standard deviation of the distribution of subgroup busy-hour traffic varies within the range of $\sqrt{y}/3.5$ to $\sqrt{y}/4.5$, if the traffic is expressed in erlangs. The exact value, of course, depends on the degree of success in load equalization, but surprisingly enough the range is rather small considering the differences in the effort which various managements devote to proper line assignments. For the present study, therefore, $\sqrt{y}/4$ appears to be suitable for the standard deviation. This relation is quite plausible. If, for instance, it is found that the variance of the distribution of a set of subgroups with small amount of traffic is $y/16$, and if the average traffic of larger groups is formed by n such subgroups each, then the variance of the distribution of such groups will be $ny/16=Y/16$ and its standard deviation $\sqrt{y}/4$. The ratio of the stand-

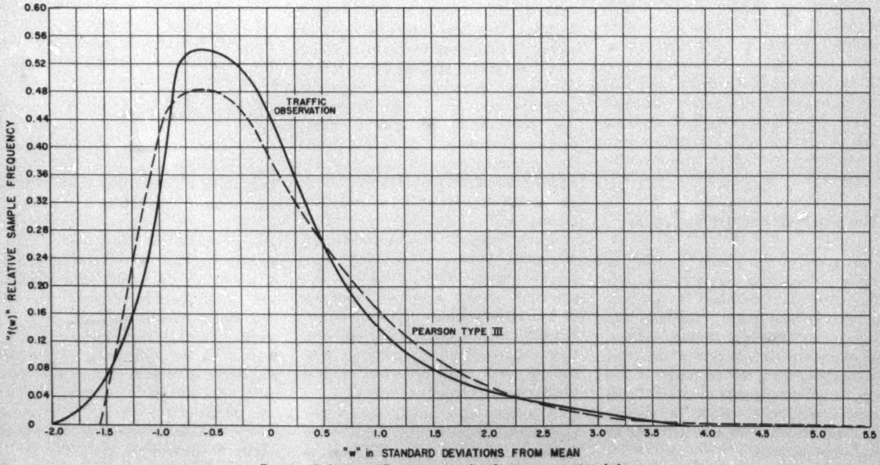


Fig. 8. Relation of Pearson-type distribution to empirical data

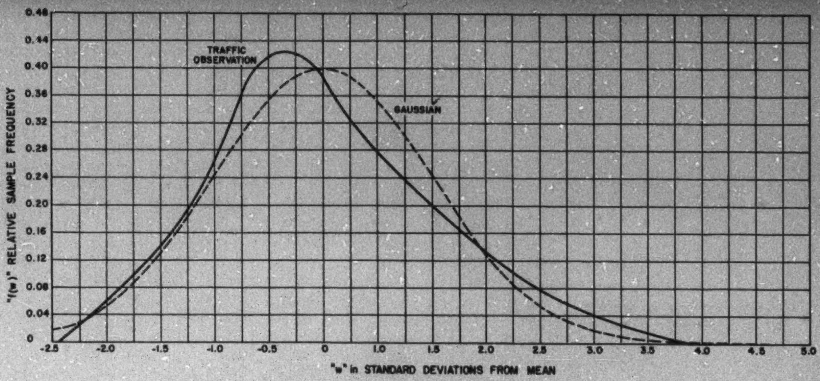


Fig. 9. Relation of Gaussian-type distribution to empirical data

ard deviation to the mean will, therefore, decrease in the ratio $1/\sqrt{y}$ with increasing average traffic. Now, very small subgroups with very small traffic have a Poisson distribution of their calling rate with a variance equal to the mean calling rate, and if this calling rate is multiplied by an average holding time of 1/20 hour, which is an appropriate value for this purpose, the standard deviation of such distribution will be $\sqrt{y}/4.5$, and the hypothesis as to combination of larger groups from such small ones follows.

Another common characteristic of all the distributions is their pronounced skewness (see Figs. 8 and 9). The distribution mode is decidedly to the left of the mean and the tails of the distribution become insignificant beyond two standard deviations to the left, but may extend over three or more standard deviations to the right. Actually one might have expected a Gaussian distribution of the subgroup traffic. If, for example, there are 95 groups of 96 lines, the traffic in each group will be the mean traffic of its lines, multiplied by the constant factor 96. According to the central limit law, no matter what the distribution of traffic over individual lines is, with increases in the number of groups, the distribution of the average traffic of the individual groups will tend to be normal. As the mean of the distribution was determined within an accuracy of ±3% at 0.95 confidence level, even a casual glance will show, and a test for goodness of fit will confirm this, that the distribution has no semblance of normality. A sample size of 95 should be sufficient to reveal normality, the traffic generated by subgroup sizes of 96 lines too small to remove the control of the population. Indeed, if the 95 subgroups are combined into 48 at random, and these into 24, etc., the range of the distribution gradually narrows, and the skewness becomes progressively smaller, until the samples and class marks are too few in number to be statistically meaningful.

Instead, when the 95 independent subgroups are organized into several ranges according to their busy-hour traffic, the number of subgroups belonging to each range form a frequency distribution. The set of subgroups can be looked upon as a sample of a hypothetical parent population, where the probabilities belonging to the various ranges are statistically reflected in the sample frequency distribution. The joint probability distribution of these range frequencies is the multinomial distribution, which with increasing number of subgroups tends to a joint normal distribution. This limiting distribution is closely related to the Pearson type III form which is also suggested by the shape of the empirical distribution.

The Pearson Type III Distribution

If the variable is standardized, this distribution is expressed in a form where skewness $g=2/M$ is the only parameter

$$f(w)=K(w+M)^{M^2-1}-Mwe^{-Mw} \tag{13}$$

Skewness g in the empirical distribution is defined as the ratio of the third moment about the mean to the cube of the standard deviation. Traffic variable w now is in units of the standard deviation with the mean at

$$w=0$$

$$M=2/g$$

$$K=M^{M^2}\exp(-M^2)/G(M^2)$$

and G is the gamma function. The distance on the abscissa between mean and mode is $-g/2$, function $f(w)$ is defined only in the range between $-M$ and infinity, and $f(w)$ is zero at $w=-M$ and asymptotic to the abscissa at infinity.

The Pearson-type distribution corresponding to the empirical data is shown dotted in Fig. 8, with the empirical skewness of 1.277. The match may not look particularly close, but it is quite good considering that at 95% confidence level the standard deviation is known only within 15%, and that there is also a limit to the accuracy of the skewness as well as in fitting the distribution to the frequency histogram. However, that the Pearson distribution is cut off at $w=-1.566$ instead of about $w=-2$, its lower modal frequency, its very long tail extending well beyond three standard deviations, and its general tendency would provide a very conservative value for the adjustment of c. In the sample frequency distribution the mean value of y was 2.572 erlangs and the standard deviation 0.382 erlang.

The cumulative probability distribution, that is, the area under the frequency-distribution curve from a point W to infinity, can be computed with the aid of a table of incomplete gamma functions as

$$\text{probability}(W<w)=\int_{W}^{\text{inf}} f(w)$$

$$=1-G_{M(W+M)}(M^2)/G(M^2) \tag{14}$$

Summing up the discussions of the last two sections, $f(w)$, as expressed by equation 13, approximates the distribution of the subgroup busy-hour traffics about their average y. Based on equation 14, the

Table I. Probability Distribution of Traffic

Traffic Variable w, from Average	Frequency, from Traffic Study	Pearson Probability	Normal Probability
to −1	0.095	0.116	0.159
−1 to −0.5	0.253	0.235	0.150
−0.5 to 0	0.252	0.223	0.191
0 to +0.5	0.179	0.152	0.191
+0.5 to +1.0	0.095	0.105	0.150
+1.0 to +1.5	0.042	0.165	0.092
+1.5 to +2.0	0.042	0.038	0.044
+2.0 to +2.5	0.010	0.025	0.017
+2.5 to +3.0	0.000	0.010	0.005
over +3.0	0.001	0.035	0.001

probability that a subgroup has a busy-hour traffic within some given range W_1 and W_2, or, in other words, the expected number of subgroups in this range relative to the total number of subgroups, can be obtained by the difference of the two corresponding incomplete gamma functions.

In Table I the deviation of subgroup traffic, in units of standard deviation, is given in the first column, the probabilities according to the traffic study and to the Pearson-type distribution as shown in Fig. 8, are given in the second and third columns respectively, and the corresponding probabilities by a normal distribution are given in the fourth column.

It may be noted that over most of the range it would be conservative to use the normal distribution than the Pearson distribution because of the higher proportion towards the above-average traffic values and lower proportion towards the lower-than-average traffic values, which would tend to bring about higher loss. Conversely, below −1 standard deviation the normal distribution shows an excess of 4%, while above +2.5 standard deviations the Pearson distribution has an excess of 3%, whereby an assumption of normal distribution would tend to reduce the loss. The Pearson distribution is more severe than the actual frequency distribution obtained from traffic measurements, in that the Pearson distribution gives greater weight to the subgroups with above-average traffic. This is also apparent from inspection of Fig. 8. In fact, the normal distribution is far more severe than the actual frequency distribution, except in the range below −1 standard deviation. The adjustment will be computed according to both the Pearson and the normal distributions, the Pearson to be used in more severe cases of expected unbalance, and the normal for applications where the load equalization is better than usual. Fig. 9 shows an empirical distribution with an average traffic of 4.964 erlangs per subgroup and 0.520-erlang standard deviation, and also a Gaussian distribution as a comparison.

The Principle of Adjustment

The loss function will be designated by $P(x, N)$, with x as the variable subgroup traffic and N the number of trunks as parameter. An attempt will be made to make $P(y+h, N+m)=P(y, N)$, with y as the average subgroup busy-hour traffic, h as an adjustment to the traffic, N the number of trunks required for y at the specified objective, and m the added trunks to obtain the same objective with the actual variation of the individual subgroup traffics. Because the adjustment is small, in the expansion of the foregoing expression all terms above the first order may be neglected. Therefore approximately the following is obtained:

$$\begin{aligned}P(y+h, N+m)-P(y, N)\\ =h\partial P(y, N)/\partial y+m\partial P(y, N)/\partial N \quad (15)\end{aligned}$$

The objective is to make the left side of equation 15 zero. This will happen if in the neighborhood of y and N the rate of increase of loss with increase of traffic is the same as the decrease of loss with the addition of trunks (and assuming fractional trunks could be added), with due consideration given to the scale factor h/m. That this is approximately true may be verified by computing the differentials, from a table of the Erlang function and its differences.

This can be simplified by analytic means. The denominator of the Erlang function $1+(x/1)+(x^2/2!)+\ldots\ldots(x^N/N!)$ can be approximated by e^{-x}, thus approximately $P(y, N)=e^{-y}y^N/N!$ which is, of course, the Poisson distribution. Then,

$$\begin{aligned}\partial P(y, N)/\partial y=P(y, N-1)-P(y, N)\\ =P(y, N)(N/y-1) \quad (16)\end{aligned}$$

The other partial derivative is

$$\begin{aligned}\partial P(y, N)/\partial N=P(y, N)(\log y-\\ (1/N!)\partial(N!)/\partial N)\end{aligned}$$

in which $\partial(N!)/\partial N$ can be obtained from the properties of the gamma functions, or by Stirling's approximation. Inserting this:

$$\begin{aligned}\partial P(y, N)/\partial N=P(y, N)(\log y-\\ 1/2N-\log N)\end{aligned}$$

the adjustment will be true if, approximately,

$$h(N/y-1)=m(1/2N+\log(N/y))$$

or, if after a series expansion of the logarithm, approximately,

$$h/m=1/2((N/y)-1)+1/(N/y))$$

The last equation can be verified numerically, for instance, by setting $m=1$ and reading out the corresponding differences h at several arguments of the Erlang tables. This usually is called the incremental trunk efficiency, which, as is well known, remains practically constant for a given grade of service when the value of change in traffic, c, is within a few per cent, and the value of traffic y is not too small.

Thus it seems justifiable to increase the average subgroup traffic y by an adjustment c, and to calculate the number of trunks with the new subgroup traffic $y^*=y+h=y(1+c/100)$ at the specified grade of service.

Adjustment c can be obtained from the requirement that the loss $P(y+h, N)$, when calculated with the original number of trunks N but with an increased average traffic $y+h$, remains the same as obtained from the actual distribution of subgroup traffics with the same number of trunks N. Therefore, first this loss $P(y+h, N)$ must be determined, then h can be found readily from the Erlang tables by extrapolating $y+h$ corresponding to N and $P(y+h, N)$.

The fundamental equation for $P(y+h, N)$ is

$$P(y+h, N)=\int_0^{\text{inf}} P(x, N)g(x)dx \qquad (18)$$

where $g(x)$, the distribution of the subgroup busy-hour traffic values, is the same as $f(w)$ from the preceding section, with x expressed in erlangs instead of the standardized variable w.

As a simplification, the Poisson form $e^{-x}x^N/N!$ is again substituted for $P(x, N)$ and thus

$$\begin{aligned}e^{-(y+h)}(y+h)^N/N!\\ =\int_0^{\text{inf}}(e^{-x}x^N/N!)g(x)dx=P(y+h, N)\end{aligned}$$

Inserting the Pearson type III distribution from the preceding section, after the change of variable w into x, and using the appropriate values for the skewness g, and $\sqrt{x}/4$ for the standard deviation,

$$P(y+h, N)=\int_q^{\text{inf}} Qx^A(x-q)^Be^{-Cx}dx \qquad (19)$$

Constants A, B, C, Q, and q can be derived from N which is given in the Erlang tables for the value y at the specified grade of service and from the constants of equation 13 by substitution. Equation 19 can again be solved with the aid of the incomplete gamma functions, after expansion of the binomial $(x-q)^B$. If B is not an integer, the variable may be changed again to $u=(x-q)$, since either

A or B or both will be an integer. Finally, h can be obtained from $P(y+h, N)$ through equation 17. The algebra and arithmetic are straightforward though lengthy for more than a few trunks N. Unfortunately the case of few trunks is of little importance.

Alternately, if the error function is used instead of the Pearson distribution for $g(x)$ a binomial expansion and a series of incomplete gamma functions will also result. The computation work in this case is less but the use of the error function is not appropriate for most applications. The approximate method to be described in the next section is simpler and its accuracy is adequate for this case.

Numerical Evaluation of the Adjustment

The starting point is again equation 18. By application of the theorem of the mean for integrals, this can be expressed as

$$P(y+h, N) = P(\bar{x}, N)\int_{0}^{\inf} g(x)dx$$

where $\bar{x}$ is the mean value of the variable. The right side can be expanded in a sum of integrals, each covering successive intervals, and each multiplied with the loss function $P(\bar{x}, N)$ at the mean value $\bar{x}$ of the variable within the interval. The intervals may range $w<-1$, -1 to -0.5, ... $+2.5$ to $+3$, $+3<w$, in units of standard deviation.

Step 1. Compute the integrals for each interval. This has been done already, in Table I, columns 3 and 4, for the Pearson type III and the normal distributions, respectively.

Step 2. Determine the mean value of the variable $\bar{w}$ for each interval, by means of the relation: $\bar{w} = \int wf(w)dw$, with both integrations carried out between the limits of the interval. The numerator was computed already under Case I for the normal distribution. For the Pearson type III distribution it will be of the form $\int w(w+q)^{B}e^{-Cw}dw$, which again can be computed by incomplete gamma functions, or, alternately, by mechanical integration. The first two steps are inherent properties of the functions, independent of traffic values and number of trunks. The computed set of values will be valid for all values of traffic and grades of service, as long as the validity of the distribution function and the numerical value of its parameter are accepted.

Step 3. Convert the standardized variable w to the traffic variable in erlangs by $\bar{x} = y + \bar{w}\sqrt{y/4}$.

Table II. Pearson Type III Distribution

Traffic y	Trunks N P=0.05	P=0.01	P=0.001	P(y+h, N) P=0.05	P=0.01	P=0.001	Adjustment c, % P=0.05	P=0.01	P=0.001
1.524	4			0.0577			5.9		
1.359		5			0.0150			12.0	
1.142			6			0.0026			23.8
2.218	5			0.0574			5.0		
2.501		7			0.0149			9.4	
2.550			9			0.0024			15.3
5.370	9			0.0564			3.2		
5.159		11			0.0140			6.0	
4.829			13			0.0023			10.8
19.99	25			0.0550			1.6		
19.49		29			0.0131			2.9	
19.74			34			0.0020			5.2

Table III. Gaussian Distribution

Traffic y	Trunks N P=0.05	P=0.01	P=0.001	P(y+h, N) P=0.05	P=0.01	P=0.001	Adjustment c, % P=0.05	P=0.01	P=0.001
1.524	4			0.0526			2.0		
1.359		5			0.0117			4.7	
1.142			6			0.0015			9.4
2.218	5			0.0524			1.6		
2.501		7			0.0116			3.3	
2.550			9			0.0014			5.4
5.370	9			0.0519			1.0		
5.159		11			0.0114			2.2	
4.829			13			0.0014			3.8
19.99	25			0.0511			0.4		
19.49		29			0.0109			1.0	
19.74			34			0.0013			1.7

Step 4. With the value of $\bar{x}$ obtain $P(\bar{x}, N)$ for each interval from a table of the Erlang function. This step needs some justification. While $\bar{x}$ is the mean value of the variable within an interval, the loss $P(\bar{x}, N)$ will be the mean value of the loss $P(x, N)$ for the interval only if its value changes linearly therein. In general, it will not be so, because then there would be no need for adjustment c. However, if the interval is narrow, say 0.5 standard deviation, the Erlang function is sufficiently linear for present purposes. As a matter of fact, at a traffic of 2.5 erlangs, for instance, 0.5 standard deviation is about 0.2 erlang, and the arguments of the tables are in steps of 0.1 erlang only. The most accurate tables are computed to six decimals, and they permit linear interpolation between steps to an accuracy of four decimals. If the grade of service is on the order of $P=0.01$, linear interpolation in steps of 0.1 erlang yields an accuracy to three significant figures. A check over several typical ranges showed that the second difference of P is nearly constant within 0.5 standard deviation of the argument, and is less than 4% of P. Therefore, the assumption that the mean value of the variable x coincides with the mean value of the loss $P(x, N)$ will be sufficiently accurate for the present purposes. At very low traffic values, about 1 erlang and below, the linear interpolation is somewhat less satisfactory, but at such values the use of Erlang's theory for traffic engineering is also objectionable for other reasons and is to be avoided if one tries to be realistic.

Step 5. Multiply each value of $P(\bar{x}, N)$ obtained in step 4 by corresponding probability $\int f(w)dw$, found in Table I, and add all these products.

Step 6. The sum obtained in step 5 is the loss $P(x+h, N)$. Enter this value into the Erlang tables at N specified for y with the objective grade of service, and find $y+h$.

Step 7. The value of added fictitious traffic will now be $h=(y+h^{*})-y$, and the adjustment $c=100h/y\%$.

Tables II and III list some typical computation results.

The values in Table II were prepared under the assumption of Pearson type III distribution for the busy-hour traffic of subgroups, similar to the one shown in Fig. 8. The adjustments are substantial, and practically the whole amount was contributed by the long tail beyond +3 standard deviations. Table III was based on normal distribution with the same standard deviation as Table II. The adjustments are more moderate here.

Adjustments to be added to the average subgroup busy-hour traffic y for the

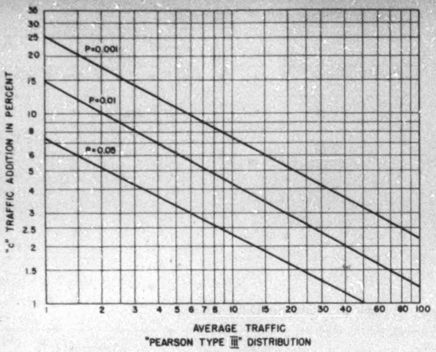


Fig. 10. Traffic adjustment for various grades of service with Pearson-type distribution

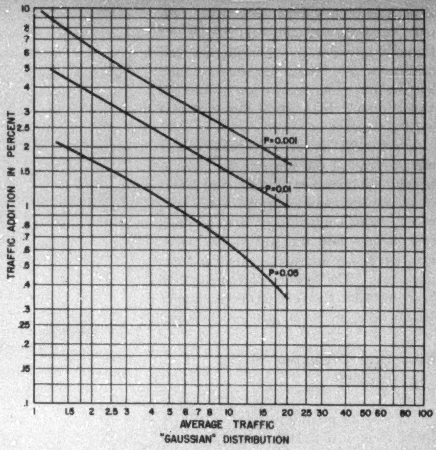


Fig. 11 (right). Traffic adjustment for various grades of service with Gaussian-type distribution

calculation of trunk quantities for various grades of service are shown in Fig. 10 and 11 for the Pearson and normal distributions, respectively. The graphs show in most parts nearly perfect linear relationship when drawn on a log-log scale. The actual tabular values reveal an almost imperceptible curvature.

Several distributions of the traffic studies were compared with the preceding theory at various average traffic values and at various grade-of-service objectives. The adjustments needed to bring the grade of service up to the objective were computed for the actual distribution, and it was found that such adjustments should be between those shown in Figs. 10 and 11, about two thirds of the way towards the Pearson-type distribution.

In conclusion, grade-of-service specification 2 prescribes that the loss of the individual subgroups in their respective busy hours, on the average, should not exceed a given amount. Then the average subgroup traffic should be increased by a certain amount *c*, and the number of trunks for the subgroups should be based on the increased average traffic. The adjustment depends on the average traffic as well as on the required grade of service, and its value can be found in Figs. 10 and 11. Fig. 10 is to be used when no attempts were made to equalize the traffic between the subgroups and Fig. 11 is to be used when special provisions assure good equalization. For the average case adjustments intermediate between Figs. 10 and 11 are suggested. When the subgroups are perfectly equalized, or if the subgroup traffics are individually known, there is, of course, no need for adjustment. The adjustment for subgroup variation is independent of and in addition to the adjustment for busy-hour phase shift. The latter was discussed under Case I, and is applied to obtain the average subgroup busy-hour traffic from the subdivided office busy-hour traffic.

Average Daily Grade of Service of Subgroups

Grade-of-service objective 3 required that the average total daily loss of all subgroups of a rank should not be more than a given amount. The exact amount of the loss is irrelevant, and can always be expressed in equivalent subgroup losses. From the traffic engineering point of view, this specification is comparable with the previous one for the computation of trunk quantities. The number of trunks is based on group busy-hour traffic, and adjustments *a* and *c* are applied as before. The only item still needed is the conversion factor between group busy-hour losses and the total-day's loss.

In Fig. 5 the decrease of traffic *b* adjacent to the busy hour is given. Assume now, as was done under Case I, that there will be two side hours with traffic y' and two side hours further removed with traffic $y''=y'(1+y'/y)/2$. There may be an additional hour with traffic comparable to y'. The remaining hours during the day will contribute negligible amounts to the total loss. This total loss is obtained by computation of the individual loss for 1 hour with y, 3 hours each with y', and 2 hours each with y'', and adding all these together. Table IV shows typical multipliers for full-availability trunking.

Two facts are apparent. The first is that for any grade of service the multiplier remains quite constant within a large range of traffic values, though it becomes smaller as traffic increases. The second fact to note is that the multiplier becomes smaller as the busy-hour grade of service improves.

Accordingly, for the most commonly used grade of service $P=0.01$ and the usual few erlangs of traffic volume found in subgroups, the total-day's loss would be about twice the busy-hour loss.

Table IV relates the traffic to the subgroup busy-hour grade of service. The objective set under specification 3 limits the average daily loss to a certain figure. It has been learned that the individual-subgroup traffics vary considerably about their average. When $N+m$ trunks were provided, *m* was determined by the desire to keep the average busy-hour loss within the objective, and for the purpose of trunk calculation the average group busy-

Table IV. Typical Multipliers for Full-Availability Trunking

Traffic, Erlangs	Total-Day Multiplier for Busy-Hour Loss		
	P=0.05	P=0.01	P=0.001
1.4	2.7	2.0	1.5
2.5	2.8	2.1	1.5
5.2	2.8	2.1	1.5
20.0	2.4	1.8	1.3
100.0	1.65	1.17	1.04

hour traffic y was increased by the amount h. Even though the average group busy-hour loss will now equal the objective, some groups will show higher-than-average, and a larger number of groups will show lower-than-average loss. According to Table IV the groups with higher-than-average loss require a larger multiplier to obtain the total-day loss while the other groups require a smaller multiplier, though otherwise each multiplier will remain constant over the range of traffic variations of these subgroups.

The over-all multiplier is designated J, and in general $J(P)$ will be a function of the loss P, and will depend on the traffic only through P. The daily average loss will be $JP(y+h,\ N+m)$, where $P(y+h,\ N+m)$ is the objective busy-hour loss $P(y,\ N)$. The daily average loss will also equal the sum of terms $J(P)P(\bar{x},\ N+m)\int g(x)dx$, and as $P(y+h,\ N+m)$ is the mean value of $P(\bar{x},\ N+m)$, J will be the mean value of $J(P)$. If the change of $J(P)$ with $P(x,\ N+m)$ is linear relative to the variable x within the range of subgroup traffic variations, then the over-all multiplier J which should be applied to the average busy-hour loss $P(y+h,\ N+m)$, will be the value of $J(P)$ corresponding to the mean value of $P(xN,\ +m)$, which, however, is just $P(y+h,\ N+m)$. This value of $J(P)$ $(y+h,\ N+m)$ can be found in Table IV. The assumptions of linear change of $J(P)$ and the propriety of the over-all multiplier were numerically verified at several typical values.

Therefore, the average grade of service provided to this rank, that is, the total-day's loss of n adjusted subgroups, will equal the average subgroup busy-hour loss $P(y+h,\ N+M)$ with average subgroup busy-hour traffic y and $N+m$ trunks in each subgroup, multiplied with the proper factor from Table IV. Whether the average busy-hour loss is the starting point, or whether the other parameters are established from the specified average total-day's loss, is irrelevant to the discussions in this section.

It is interesting to note that if adjustment c is not applied to the average traffic y, but only the number of trunks N corresponding to the average traffic are provided, then the daily-loss multipliers for the different busy-hour grade-of-service objectives will be much closer to each other than those listed in Table IV. The reason for this is that while the multipliers increase with increasing busy-hour loss, the amount $[P(y+h,\ N)-P(y,\ N)]/P(y,\ N)$ decreases with increasing busy-hour loss, and the two tendencies will compensate each other to some degree. For the lower traffic values the multiplier will vary between 2.9 and 3.9 (3.9 for $P=0.001$), and for about $y=20$ erlangs between 2.1 and 2.6 (2.1 for $P=0.001$).

For graded trunk groups the multiplier has somewhat different values. At $P=0.01$ it is about 2.3 to 2.4 between 10 and 25 erlangs.

Office Busy-Hour Grade of Service

Grade-of-service specification 4 requires that the loss during the office busy hour should not exceed the objective. Considering again a certain rank, for example, all connectors in the office, the combined office busy-hour loss of the connectors is the sum of the losses of the individual connector groups during the office busy hour.

The average traffic of a connector group is Y/n, where Y is the total office busy-hour traffic of the connector rank, and n the number of connector groups, as before. This traffic will also show variation from group to group, as did the group busy-hour traffic y before, but, in addition, this variation will be superimposed on a variable mean value between the limits y and y' due to the phase shift of the group busy hours. The mean traffic of the latter variable will be, of course, Y/n, and the factor A from Case I is a measure whether Y/n lies closer to y or to y'. Examination of Figs. 5 and 6 reveals that at about 2 erlangs Y/n is about halfway between y and y'; below this it is closer to y', and above closer to y. Above about 20 erlangs it is practically wholly composed of y.

In Case I, the conclusion was reached under the assumptions made there, that the prorated average subgroup traffic Y/n during the office busy hour may be considered as composed of two components: 1. Ay, from those subgroups whose subgroup busy hour hypothetically coincides exactly with the office busy hour; and 2. $(1-A)y'$, from the subgroups which with their side hour completely cover the office busy hour. Adjustment c of the first component is to be computed by insertion of y into the equation for $P(y+h,\ N+m)$, and adjustment of the second component is to be computed by insertion of y' into the equation $P(y'+h',\ N+m)$.

If the change of loss function $P(x,\ N)$ were linear within the range between y and y' with change of x, that is, $P(x_1,\ N)/P(x_2,\ N)=x_1/x_2$, the loss at Y/n would be $P(Y/n)+h^*,\ N+m)$, where h^* is the traffic increase for Y/n, and adjustment $c(Y/n)$ could be read from Figs. 10 and 11 at abscissa Y/n. As can be seen from Fig. 5, the distance between y and y' is about 35% at 1 erlang, 26% at 2 erlangs, and 18% at 5 erlangs. At low traffic values, say below 2 erlangs, and particularly below 0.01 grade of service, the second differences of the loss functions are not negligibly small within the range y to y', so a better approximation is required.

In the following, $Y/n=y^*$ will be designated y^*. The traffic adjustment per subgroup is again h, corresponding to average subgroup busy-hour traffic y, and the adjustment corresponding to the office busy-hour average traffic y^* will be h^*, which can be obtained from Figs. 10 and 11. It can be suspected, however, that, due to the subgroup busy-hour phase shift, this adjustment must be increased, say to h^*+q, and the adjusted actual number of trunks N (which were formerly designated $N+m$) may be then found from the Erlang tables at the argument y^*+h^*+q.

The average office busy-hour loss will be composed of the sum of elementary losses, similar to the integrand of equation 18, say $P(x,\ N)g(x)dx$, with the conditional probability that the mean value of the variable subgroup traffic x of that particular subgroup during the office busy hour is z. The unconditional probability for the latter is $f(z)dz$. As the two distributions, $g(x)$ for the variation of subgroup traffics because of inadequate equalization and $f(z)$ for the phase shift of subgroup busy hours, are independent of each other, then

$$P(y^*+h^*+q,\ N) = \int_x \int_z P(x,\ N)g(x)f(z)dzdx \quad (20)$$

The limits of integration do not need to be defined explicitly for the moment; in general they will cover the whole range of possible variations of x, and z will be limited between y and y'. The order of integration is immaterial in the present case. $\int P(x,\ N)gx\ dx=P(z+h,\ N)$ from equation 18. Now $P(y^*+h^*+q,\ N)$ is expanded into a Taylor series:

$$P(y^*+h^*+q,\ N)=P(y^*+h^*,\ N)+ qdP(y^*+h^*,\ N)/d(y^*+h^*)+\ldots$$

The second derivative of P is fairly small, even at low traffic values and $P=0.001$, and q is very small, perhaps a few per cent of P. Therefore, second- and higher-order terms may be omitted.

From equation 16

$$dP(y^*+h^*,\ N)/d(y^*+h^*) = P(y^*+h^*,\ N)(N/(y^*+h^*)-1)$$

and

$$P(y^*+h^*+q,\ N)=P(y^*+h^*,\ N)\times [1+q(N/(y^*+h^*)-1)]$$

Furthermore, $z+h=y^*+h^*+(z-y^*)+j$ are set where $z-y^*$ is the variation of subgroup traffic about Y/n during the office busy hour, and $j=h-h^*$ is the variation of the adjustment h about h^*. Expanding again, and with the same approximation as before

$$P(z+h,\ N)=P(y^*+h^*,\ N)+(z-y^*+j)\ dP(y^*+h^*,\ N)/d(y^*+h^*)+\ldots$$

$$=P(y^*+h^*,\ N)[(1+(z-y^*+j)\times(N/(y^*+h^*)-1)]$$

Combining the preceding equations and after some simplifications

$$1+q(N/(y^*+h^*)-1)=\int 1+(z-y^*+j)\times(N/(y^*+h^*)-1)f(z)dz$$

The integral $\int f(z)dz=1$ if extended over the whole range of possible values of z, by definition. Thus

$$q=\int(z-y^*+j)f(z)dz$$

The integral $\int(z-y^*)f(z)dz=0$ since y^* is the mean value of z within the range also be definition. Finally

$$q=\int jf(z)dz \qquad (21)$$

Observing that adjustment $c=100\ h/y$, and that Fig. 10 suggests the form $c=J/y^K$, the following empirical approximate relations are obtained:

$$\left.\begin{array}{l} h=0.25\ \sqrt{y} \text{ for } P=0.001 \\ h=0.15\ \sqrt{y} \text{ for } P=0.01 \\ h=0.075\sqrt{y} \text{ for } P=0.05 \end{array}\right\} \qquad (22)$$

It is interesting to note that for $P=0.001$ the adjustment is the same as the empirical standard deviation for $g(x)$ and that the adjustment is about halved in each of the foregoing steps between grades of service. Now, $j=h-h^*=J(\sqrt{z}-\sqrt{y^*})=J\sqrt{y^*}(\sqrt{(z/y^*)}-1)$, where J is one of the constants in equation 22.

Now consider $f(z)$ and the limits of integration. $f(z)$ is the frequency distribution of subgroups with traffic z during the office busy hour. In Case I it was shown that the office busy-hour traffic z varies between y and y' according to $y-t_0(y-y')$ (only absolute values of t_0 taken) between $-1<t_0<+1$, and stays y' beyond. t_0 was the center point of the group busy hour, the center point of the office busy hour being 0.

The distribution of t_0 was assumed to be normal with $1/T$ standard deviation. If the foregoing limits are disregarded for the moment, it may be said that the distribution of z will be normal, with $(y-y')/T$ as standard deviation.

By a change of variable $u=T(y-z)/(y-y')$ the distribution $p(u)$ will be normal with unit standard deviation. $u=0$ will be at $z=y$. Only values $u>0$ will be considered, and as t_0 may take positive as well as negative values yielding the same z, the frequency for each interval (u to $u+du$) will be $2p(u)du$. The lower limit of integration is at $u=0$. The normal distribution is incomplete and stops at T standard deviations, where the variable $z=y'$.

Beyond $u=T$, the variable $z=y'$ over the whole tail of the normal distribution.

Equation 21 will now read:

$$2T/(y-y')\int jp(u)du$$

and

$$j=J\sqrt{y}[(1-u(1-y'/y)/T)^{1/2}-\sqrt{(y^*/y)}]$$

Under Case I, $(1-y'/y)$ was $b/100$, the decrease of traffic from the busy hour to the adjacent hour (relative to y), or b if expressed as a fraction instead of as a percentage. The limit of integration is T, and what is beyond is small, so the term ua_1/T, in general, will be less than 1, on the order of 0.35 or less according to Fig. 5. Thus

$$(1-ub/T)^{1/2}=1-ub/2T$$

approximately. The error of this approximation is on the order of 0.01 in the worst case of practical significance. It decreases rapidly with increasing traffic and, furthermore, it is on the conservative side. Therefore,

$$j=J(\sqrt{y}-\sqrt{y^*})-Jub\sqrt{y}/2T$$

approximately. Again, $y^*/y=1-a_1$ from Case I, if a_1 (Fig. 7) is expressed as a fraction instead of as a percentage, a_1 is the adjustment for the busy-hour phase shift, which is subtracted from y to obtain $Y/n=y^*$. Finally

$$j=J\sqrt{y}(1-\sqrt{(1-a_1)}-ub/2T) \qquad (23)$$

also, from equation 21,

$$q=2T/(y-y')\left(\int_0^T jp(u)du+ j(u=T)\int_T^{\text{inf}} p(u)du\right) \qquad (24)$$

After some rearrangements, keeping in mind that

$$2\int_0^{\text{inf}} p(u)du=1$$

$$q=[2TJ\sqrt{y}/(y-y')][1-\sqrt{(1-a_1)}]-(2Jb\sqrt{y}/(y-y')\left(\int_0^T up(u)du+ T\int_T^{\text{inf}} p(u)du\right)$$

$$=[2J\sqrt{y}/(y-y')]\left\{T[1-\sqrt{(1-a_1)}]- b[p(0)-p(T)]-bT\int_T^{\text{inf}} p(u)du\right\} \qquad (25)$$

As before, $p(u)$ is the tabulated error function and $\int p(u)du$ the tabulated error integral.

At grade of service 0.001, the adjustment values from equation 25 are 2.16%, 1.01%, and 0.73% at Y/n 1, 2, and 5 erlangs, respectively. For a grade of service 0.01 or 0.05, the adjustments are about one half or one fourth of these, corresponding to the values of J in equation 22. This fictitious incease is over and above the increase c due to subgroup traffic variation, and supposed to be used in determining the number of trunks per subgroup from the prorated office busy-hour traffic Y/n, to protect the specified average grade of service during the office busy hour. Now, from Fig. 10 the adjustments c corresponding to the foregoing traffics at 0.001, were 25%, 17.5%, and 11%; and it will be recalled that there was considerable latitude in the choice of intermediate values between Figs. 10 and 11, depending on the degree of traffic equalization in a particular application. The magnitude of q is probably less than the error in judgment in the choice of intermediate values between Figs. 10 and 11, therefore (in case of grade-of-service specification 4) if in the choice of c Fig. 10 is used, adjustment q may be disregarded.

A few words may be in order on the possible errors in these discussions. In addition to the assumptions listed under Case I, in several series expansions second- and higher-order terms have been neglected, and also the Poisson function substituted for the Erlang function. Other possible sources of error can be found in linear interpolation of tabulated functions. In addition to these analytic errors, there were others made when the distribution functions for the subgroup traffic variation, side-hour traffic decline, and busy-hour phase shift, as well as their parameters (mean, standard deviation, and skewness) were determined from a limited amount of empirical information. Compared with these statistical uncertainties, as expressed by the latitude given between Figs. 10 and 11, the analytic errors are trivial. But even including the statistical errors the various adjustments appear to be close enough, if the limited accuracy of traffic information available for practical applications, the vagueness in what can be considered good grade of service, and the degree to which the various theoretic loss formulas come close to the facts of life are considered.

Superficially, it may seem that the discussions related to grade-of-service objectives 3 and 4 are contradictory. Consider the example of an office with 10,000 stations originating 250 erlangs busy-hour traffic. The stations are divided into 100 groups with 100 stations each, with a prorated average traffic Y/n of 2.5 erlangs per groups. If specification 2 is followed, say at the grade of service 0.01, adjustment $a=12.5\%$ from Fig. 6 must first be applied, so that the group busy-hour traffic will be 2.82 erlangs. Each group will lose 0.0282 erlang in its busy hour, or 2.82 erlangs for all 100 groups. The total-day's loss of all 100 groups will be, according to Table IV and specification 3, $2.1\times 2.82=5.92$ erlangs, which is $2.35\times 0.01\times 250$ erlangs. To obtain the number of trunks per group, add to 2.82 erlangs a second adjustment c, which is 2.4% from Fig. 10, so $y+h=2.9$ erlangs, which require 7.8 trunks.

If specification 4 is followed, the average traffic per subgroup will be 2.5 erlangs; $c=2.6\%$ is added and 7.2 trunks per group specified for 2.57 erlangs. It may be reasoned that since the trunks were computed for the office busy hour for a grade of service 0.01, that is, with a loss of $0.01\times 2.50=0.025$ erlang, the total-day's loss for the rank according to Table IV should be $100\times 2.1\times 0.025=5.25$ erlangs, which is 13% less than obtained under specifications 3 with about 8% more trunks.

Actually, no matter which way the number of trunks is figured, each group will still carry in its busy hour on the average the same 2.82 erlangs traffic, and the group busy-hour grade of service now, with only 7.2 trunks, will be 0.0166 instead of 0.01 with 7.8 trunks. The total-day's loss is perhaps $2.2\times 0.0166\times$ 2.82 erlangs for each group, and 100 times as much for 100 groups, or 10.3 erlangs altogether, or $4.13\times 0.01\times 250$ erlangs. This is 75% more than with 7.8 trunks per group and nearly twice as much as was suggested earlier in this paragraph.

The fallacy of the reasoning is that it did not take into consideration the fact that objective 4 is concerned only with the grade of service during the office busy hour, and is satisfied that, in general, during the rest of the day, the service is better. Objectives 3 and 4 represent completely different concepts, and there is no simple connection between them other than the relationship revealed by analysis in this paper.

The Office Grade of Service

The preceding discussions were all concerned with measures for sustaining the grade-of-service objective in a single switching rank. However, a telephone office contains several switching ranks, and some consideration must be given to the over-all grade of service provided by the office. For example, consider a central office consisting of a primary rank, four ranks of selectors, and one rank of connectors; and again, the total office busy-hour traffic is 250 erlangs. The subgrouping in the primary rank is extensive, in groups of 100 or 200 lines, and, likewise, the connectors are arranged for 100 terminals in a group. The selector rank preceding the connectors consists of groups carrying traffic to 1,000 terminals. The two sets of trunks between the remaining three selector ranks carry the large volume of intraoffice traffic as well as traffic to other switching offices. In addition, there will be trunks to the third selector rank which receive traffic from other switching offices.

To begin with the interoffice trunking, there is no generalization possible about the outgoing and incoming traffic volumes relative to the total, nor as to the location of the group busy hours relative to the office busy hour, because these will depend on peculiarities of local situations. Therefore, the previous rules for subdivision, in general, will not be applicable, since they refer to a more or less random subdivision. Instead, the traffic in each trunk group should be appraised from whatever individually pertinent information is available. On the other hand, Figs. 6 and 7 are still useful to express the relationship between the group and office busy-hour traffics on such trunk groups, in the absence of more definite information.

Assume now that the total traffic leaving the office and the total traffic received from outside are equal and that their over-all busy hours coincide with the office busy hour. These assumptions simplify the problem and are very nearly true in many situations. Where the assumptions do not fit, the general principles developed in these discussions should enable modifications to be applied. The intraoffice traffic is carried in one large volume and requires no adjustments for subdivision.

Due to the fact that switching does not commence until dial tone is received, the number of trunks per group in a primary rank is computed differently from the methods used in the previous discussions. In spite of these differences, whether the grade of service is based on calls being lost or delayed, the consequences of subdivision are similar, and so to simplify matters further it is assumed that the trunking in the primary rank is based on grade-of-service specification 2 or 3, with the adjustments for subdivision carried out accordingly. The adjustments for the connector rank, that is, for the computation of the trunks from the selectors preceding the connectors, were discussed fully in previous sections. The trunks preceding the last rank of selectors are divided into ten groups carrying on the average 25 erlangs traffic in the office busy hour. The two adjustments, one for busy-hour phase shift and the other for group variation, are again based on the previous principles, with 25 erlangs average group traffic.

The end result for the whole office, however, will be something unexpected. Take a call originated during the office busy hour in one of the primary subgroups and assume objective 2 with a loss of 0.01 during the subgroup busy hour. If the subgroup busy hour coincides with the office busy hour, 1% of the attempts will be lost, and similarly in those ranks where the trunking was based on the office busy hour. In other ranks, for instance, interoffice trunks based on group busy hours, as well as in the connector rank and the selectors preceding it, and computed to objective 2, during the office busy hour less than 1% of the attempts are lost on the average, so the grade of service received will be better than the sum of the grade-of-service objectives over all ranks. If the busy hour of the primary group was away from the office busy hour, the grade of service will be still better, because, on the average such calls will terminate in groups having their side hours at that time.

Now, objective 2 attempted to provide an average grade of service to the aggregate of subgroups, and to equalize the effect of the busy-hour phase shift. But the end result is that while the groups with their busy hour during the office busy hour will suffer a slightly higher loss than the others, all the groups will have less loss than 0.01 times the number of switching ranks. This is not necessarily undesirable, nor in conflict with specification 2, but it could hardly be anticipated from the stated objective.

If specification 4 is followed by application of adjustment h^*, or more precisely h^*+q, to Y/n, only enough trunks are provided to limit the loss in each rank and group during the office busy hour to, say, 0.01. A call originated during the office busy hour will then have 0.01 loss through every rank which has not been subdivided, in accordance with assumptions made that the same total volume of traffic is carried during the office busy hour through every rank. In those ranks which are subdivided, groups which have their busy hour during the office busy

Table V. Comparison of Objectives 3 and 4 Losses, in Erlangs

Trunks	Objective 3 Loss		Objective 4 Loss	
	Group Busy Hour	All Day	Group Busy Hour	All Day
Primary	2.82	5.92	2.50	10.3
First to second	2.50	6.00	2.50	6.0
Second to third	2.50	6.00	2.50	6.0
Third to fourth	2.55	5.35	2.50	6.8
Fourth to connectors	2.82	5.92	2.50	10.3
Total		29.19 (11.7% of 250-erlang office busy-hour traffic)		39.4 (15.8% of 250-erlang office busy-hour traffic)

hour will suffer a higher loss, and the other groups a lesser loss, but even in these ranks the loss will be equal on the average to the objective. Therefore, it appears that specification 4 can be more consistently administered through all ranks in the office, but also is more inconsistent in the grade of service between subgroups originating the traffic. As was mentioned before, the two objectives represent entirely different concepts for the grade of service: objective 2 protects an average grade of service, while objective 4 insures a better than minimum quality throughout the day.

These views can be extended to the whole day. It has already been learned that, for trunk engineering, objective 3 is equivalent to 2. But surprisingly, although it was just shown that all traffic originated in the subgroups during their group busy hour will have an over-all loss less than the anticipated objective, the total loss over all ranks and throughout the day will be almost exactly equal to the objective, or perhaps a trifle more. This will be apparent from Table V, and the reason for it is that the loss over all the ranks is spread over several side hours of the originating subgroups.

No specification equivalent to 3 can be found for 4 since 4 is inherently unconcerned with what happens outside the office busy hour. In view of Fig. 6 and Table IV, the relationship between objective 4 and total-day's loss must be highly dependent on the traffic and grade of service, which would be hardly worthwhile to develop. In the present example, the 10,000-station office with 250 erlangs busy-hour traffic, this comparison of objectives 3 and 4 can be made.

In Table V it was assumed that the primary groups consist of 100 lines each, that the same traffic is carried through each rank, and that the trunking between first and second, and second and third selectors were the same for both objectives, with an all-day multiplier of about 2.4 applicable to graded trunk groups. The total-day's loss of the whole office under specification 4 is of course higher by about 35%. With specification 4 the saving in equipment in the primary and connector ranks is about 8%, and in the rank preceding the connectors about 1.5%. The percentages lost as shown in the table refer to the office busy-hour traffic. If this is related to the total-day's traffic, assuming that the busy-hour contains about one tenth of the 2,500-erlang total daily traffic, the day's loss would be about 1.2% and 1.6% of the daily traffic, with specifications 3 and 4, respectively. From a practical point of view there is little noticeable difference between the two: objective 4 is satisfied with less equipment and is easier to administer because adjustment *a* need not be applied, but objective 3 is somewhat more equitable.

In addition to those discussed before, two more grade-of-service specifications may be worth considering.

5. For a central office, the grade of service given the combined traffic during a specified period, say over 24 hours of the average business day, should be equal in each rank, and should not be worse than the specified amount. This specification would be even more complex to administer than 3, because the methods of objective 3 or 4 would have to be applied to each rank with a busy-hour grade of service which anticipates the total-day multipliers. There would be nothing gained except some superficial symmetry. A more efficient objective is given in specification 6.

6. For a central office, the over-all grade of service given to the combined traffic of all ranks during a specified period say over 24 hours of the average business day, should not be worse than the specified amount, with the additional provision that the total cost of the equipment should be minimum.

Objective 6 is obviously the most efficient of all those discussed, but its administration is formidable. Not only must each rank be computed with a grade of service appropriate to the relative cost of the type of equipment employed in the rank, but to minimize the total cost the grade of service must also be modified according to the trunking method and the traffic in the rank. In practice this objective is being considered to a limited extent by assigning different grades of service to different ranks, which in a broad manner takes into account the relative costs of the types of switches.

It was already said that the guiding principles for the grade-of-service objectives are obscure, and that their theoretic measures are just broad approximations, based on a number of simplified assumptions, such as statistical equilibrium, Poisson distribution of attempts, disposition of calls during congestion, etc., which in many actual applications are not even approximately correct. Then what is the purpose of such grade-of-service objectives? There are three important justifications in addition to the fact that in surprisingly many cases the theoretic predications are borne out by traffic measurements, in spite of the unreality of the assumptions. First, experience has shown that if certain grade-of-service rules are followed, the day-by-day service results are very satisfactory to telephone users. Second, it provides an exact theoretic yardstick to compare the performance of widely different systems or arrangements. Third, and perhaps most important, a good grade of service protects the service from collapse during unexpected overloads by providing spare capacity for such eventualities. To see this, consider a traffic of 10 erlangs. At the grade of service 0.1 about 15 trunks are provided. If temporarily the traffic were to rise 50%, the trunks would be 100% loaded and unable to carry any additional traffic. On the other hand, on the basis of 0.01 19 trunks would have been provided, and they could carry almost 100% overload before the same catastrophic condition could occur.

Conclusions

In the preceding discussion various adjustments were developed, which are to be applied to average traffic values when large traffic volumes are subdivided and the number of trunks is determined for each subdivision with certain grade-of-service objectives to be satisfied. In general, the need for these adjustments arises because the busy hours of the subdivisions are shifted in phase relative to each other, and because in most cases it is impractical completely to equalize the traffic volumes between the subdivisions. There is a separate additive adjustment for each of the two causes; the first may be omitted if only the over-all busy-hour grade of service is of interest, and the second may be omitted if the traffic between the subdivisions is well equalized. The first adjustment may also be used to determine the contribution of a trunk group to the over-all busy hour if its group busy-hour traffic is known or vice versa.

The relations between the busy-hour grade of service and the total-day's grade of service were also developed. The grade of service to be provided by switching equipment can be prescribed from different points of view; in general these either attempt to protect the average grade of service given to the different subdivisions, or insure that the over-all service should not be degraded below a certain level during any one hour of the day. The most efficient objective minimizes the total cost of the equipment within the permissible grade of service. Its administration is very complex if an attempt is made to carry it to perfection.

The latitudes in the application of these adjustments are necessarily broad, consistent with the limitations inherent in all traffic engineering problems. These limitations include: diverse conditions in individual applications which prevent generalization, lack of pertinent data for such applications, practical limitations in control of fluctuations, vagueness in basic principles for grade of service, and deviations of actual conditions from the various simplified assumptions which are fundamental to the congestion theories.

Pagemaster Receiver and Modulation Equipment

J. W. YOUNG
NONMEMBER AIEE

AT PRESENT, there are many varied types of paging systems on the market which are applicable to the many types of services or industries requiring paging systems. Each of these, whether they be audio, inductive loop, r-f (radio-frequency), light, bell, or any other type of system, has certain advantages and limitations. The results of a comprehensive investigation of the paging field showed that there was a definite need and a market for a selective r-f paging system. In many instances such a system could adequately replace, with improved operation, the existing type of system. In other instances it could advantageously supplement an existing paging system with a resultant improvement in the over-all paging operation.

The results of the aforementioned study led to the development of a practical type of selective r-f paging system known as the Pagemaster* system. The Pagemaster system is a relatively new type of communication system whereby one of several thousand personal pocket-carried receivers may be selectively signalled. In essence, the system is a one-way signalling system in that no communication is initiated by the pocket-carried receiver. The receiver is essentially a signal-actuated device whereby contact is established but not maintained. The receiver will indicate that contact has been established by emitting an audible tone.

The emphasis to date has been placed on an industrial type of system which can accommodate 455 receivers. This industrial system has found wide acceptance in various industries and hospitals where 455 receivers would be more than adequate. However, the same principles used in the industrial system can be extended into a system capable of servicing 3,235 individual subscribers, and would therefore be adaptable to city-wide paging systems.

This paper will describe the coding and modulation system employed in this r-f paging system; the base station equipment required to code the information; and the receiver and its decoder.

System Operation

Before the modulation system and the components of the system, it might be advantageous to describe briefly its over-all operation. The city-wide paging system's operation is given in a companion paper.[1] An acquaintance with the complete operation will facilitate an understanding of the functioning of each of the components used to achieve the end result.

The complete system or installation is comprised of an encoder, a transmitter, an antenna, and any number of receivers. Since the system is selective, each of the receivers will have a number which will correspond to a predetermined code. In the system a total of nine different digits are used and, as will be explained later, each digit will correspond to a different audio frequency.

When it is desired to contact a person assigned a receiver, the calling party will contact the operator who has control of the base station equipment. The operator in turn will set the desired code into the encoder and place the equipment in operation. The output from the encoder which will be the coded modulating signal will amplitude-modulate the transmitter, and the desired code will be radiated from the antenna. All of the receivers in the system will receive this coded signal, but only the receiver with the proper code will respond, by emitting an audible tone. The called party will merely press a button on the receiver which will squelch the tone and reset the receiver automatically for the next call. Upon being paged, the called party will take some predetermined action, which is usually to call his office or some other designated telephone.

Coding and Modulation System

The nine different audio tones employed in the coding system range in frequency from 160 cps (cycles per second), which corresponds to digit 1 to 301.18 cps, which corresponds to digit nine. The frequencies corresponding to each of the digits are shown in Table I. The choice of the audio frequencies was dictated by the decoding mechanism in the receiver which utilizes tuned vibrating reeds, to be discussed in connection with the receiver.

The coded signal will consist of sequential bursts of the appropriate audio frequencies. Fig. 1 shows the modulating signals employed in the 3-digit and 4-digit systems. In the 3-digit system, which is synonymous with a 3-reed system, the transmitted signal consists of a carrier

---

Paper 58-814, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted April 1, 1958; made available for printing September 17, 1958.

J. W. YOUNG is with the Stromberg-Carlson Company, Rochester, N. Y.

* Registered trade-mark.

Table I. Frequencies Relating to Digits

Tone	Frequency, CPS	Division Factor	Feedback Applied, Stage No. 2	3	4	5
1	160.00	128	no feedback			
2	170.67	120		X		
3	182.86	112			X	
4	196.92	104		X	X	
5	213.33	96				X
6	232.73	88		X		X
7	256.00	80			X	X
8	284.44	72		X	X	X
9	301.18	68	X	X	X	X

amplitude modulated with three sequential bursts of audio frequency. Each of the bursts are 0.5 second in duration with a gap of approximately 0.15 second between the bursts. The coded signal is therefore 1.8 seconds in duration with a guard time of 3.2 seconds between code groups. As an example, assume that receiver 179 is to be paged. The coded signal would consist of a 0.5-second burst of tone 1 (160 cps) followed by a 0.15-second gap; a 0.5-second bust of tone 7 (256 cps) followed by a 0.15-second gap; and a 0.5-second burst of tone 9 (301.18 cps). There would be a time interval of 3.2 seconds before the next code was transmitted.

The coded signal for a 4-digit, or 4-reed, system is made up in a similar manner. In this case four bursts of tones are transmitted. In the 4-reed system the length of the tone duration, the gap interval, and the guard band are somewhat different from those of the 3-reed system. The 4-reed system is more likely to be used in city-wide paging systems where traffic may be a problem. To ensure the maximum amount of traffic within the capabilities of the system, the time required to transmit a code group is held to a minimum. In this case the coded signal is 1.4 seconds in duration with a guard time of 2.5 seconds between code groups. With this timing, approximately 15 codes per minute can be transmitted. In the 3-reed system a maximum of 12 codes per minute can be transmitted.

As mentioned previously, the choice of the audio frequencies was dictated by the decoding mechanism in the receiver. A similar situation exists in determining the time intervals in the coded signal, namely, the timing is dictated by the decoder. This will also be discussed in connection with the receiver.

Encoder

The encoder is the heart of the system as it controls the transmitter modulation, generates the proper tones, selects the tones, and transmits them in the proper sequence to the transmitter. A block diagram of the encoder used in the 3-digit system is shown in Fig. 2.

TONE GENERATOR

One of the functions of the encoder is to generate the proper tone frequencies. This is accomplished in the tone generator section of the encoder in the following manner.

The first stage is a crystal-controlled audio-oscillator which operates at a frequency of 20.48 kc ±0.012%. The output from the oscillator, which is sinusoidal, is coupled to the second stage, where the signal is shaped to provide a pulse that has a sharp negative edge. There will be one negative pulse out of the shaper stage for each cycle of oscillator voltage input. The frequency is still 20.48 kc.

The next stages (1 through 7) are identical binary dividers, a common type of plate-coupled bistable multivibrators. Because they are binary dividers, it takes two input pulses to provide one output pulse. With seven of these stages in cascade the dividing chain will divide by two, raised to the seventh power or 128. Thus for an input frequency of 20.48 kc to the first binary divider; the output frequency of the last divider will be 160 cps, which corresponds to the frequency of tone 1.

When the tone 1 frequency has been generated, the next problem is to generate the other eight frequencies or tones. This is accomplished by applying feedback to the proper stage or stages in the binary divider chain. The action of channeling feedback pulses from the output of the dividing chain into any one or more of the stages in the dividing chain is to alter the division factor of the dividing chain.

To develop the frequencies required for tones 2 through 9, feedback is applied to as many as four stages simultaneously. Considering the first five stages of the dividing chain, the division factor with no feedback will be 32. With feedback applied to as many as four stages, any division factor between 17 and 32 can be obtained. The over-all division factor can be determined by multiplying this division factor by four. Table I shows the required over-all division factors necessary to generate the required nine tones and a list of the stages to which feedback must be applied to generate these tones.

The output from the last divider will then be any of the nine tones, depending on the division factor. This output signal is applied to the output stage which in turn feeds the signal to the modulator stage in the transmitter.

The reset (feedback) pulse is generated by taking an output from the sixth binary stage and coupling it into a reset pulse generator. The reset pulse generator differentiates and amplifies the input

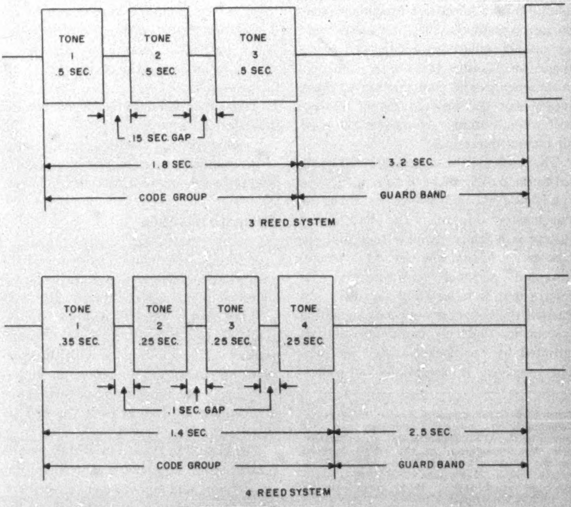


Fig. 1. Modulating signals employed in 3-digit and 4-digit systems

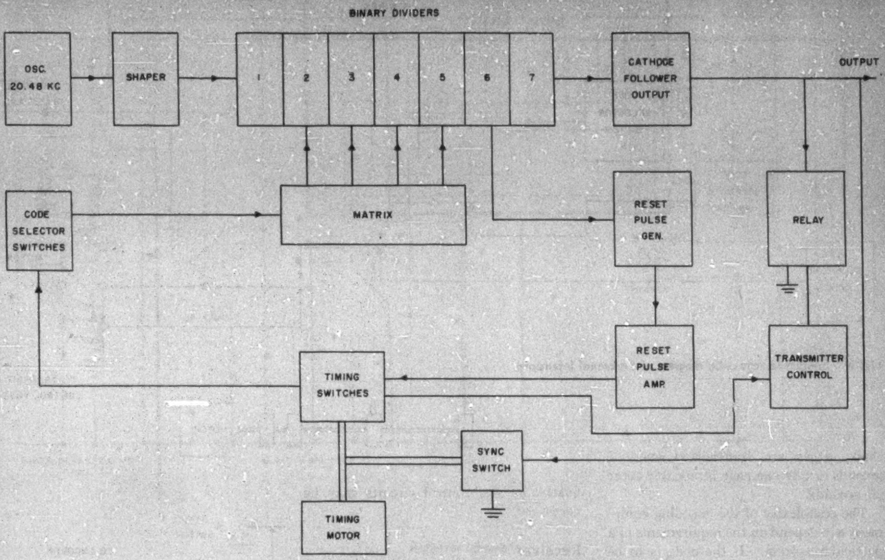


Fig. 2. Encoder block diagram used in 3-digit system

pulse. The output from the reset pulse generator, which is a sharp negative pulse, is further amplified in the reset pulse amplifier stage and channeled through the interrupter, the code selector switches, and the neon bulb matrix, to apply feedback to the proper binary stages.

The reset pulse amplifier was necessary to assure that the reset pulse would be of sufficient amplitude to apply feedback to all four stages when tone 9 was required. The position of the code selector switches determines to which stage the feedback is applied, and therefore determines the frequency of the tone generated. The matrix converts the settings of the code selector switches, which are in decimal notation, into an equivalent binary form. Neon bulbs are employed in the matrix since they are essentially open circuits unless ionized. The ionization time delays the reset pulse the required amount of time.

INTERRUPTER

The tones are sequenced properly with the aid of a timing motor and a combination of switches. The complete assembly is referred to as an interrupter. A simplified schematic diagram of a 1-channel interrupter is shown in Fig. 3.

The timing motor drives a shaft on which are mounted three cams which activate three spring-type transfer switches. The three switches provide a path to the code selector switches for the reset pulses. The settings of the code selector switches will determine the frequency of the tones transmitted.

After the code of the receiver to be paged is set up on selector switches, the code bank switch is turned on. This starts the interrupter motor and cams rotating. In rotating, the cams will activate their associated switches. The design and adjustments of the cams are such that the first switch will be on the off-normal position for the full duration of the code group. The second switch will be in the off-normal position for the last two thirds of the code group time, and the third switch will be in the off-normal position for the last third of the code group time. Thus, in operation the reset pulse will be channeled through the first, second, and third code selector switches in that order as the cams rotate. This action will result in three sequential tones being generated.

The gaps are inserted by utilizing an additional cam and switch. This switch has a single set of contacts which, in the normally closed position, ground the output of the encoder. The cam has three successive notches and, when the arm of the switch falls into one of the notches, the ground on the output of the encoder is removed. By properly synchronizing all of the cams the output of the encoder for a 3-digit system will be as shown in Fig. 1.

The encoder also provides the necessary control of the transmitter. When the first switch in the interrupter is in the off-normal position, the grid of the transmitter control tube is removed from ground and the tube is allowed to conduct. This tube controls a relay which has several sets of contacts: one set controls the transmitter carrier; an additional set controls the output from the encoder. The transmitter carrier will be on only when a code group is being transmitted. As long as the code bank switch in the encoder is in the ON position the timing motor will continue to rotate and the code will be transmitted once for each revolution of the cams.

ENCODER VARIATIONS

The operation of a single-channel encoder has been described. In the encoder for an industrial system four channels are available. This requires that there be four sets of code selector switches and four sets of switches in the interrupter circuitry. The timing motor runs at a speed of 3 rpm. Since the cam shaft is direct-driven from the motor

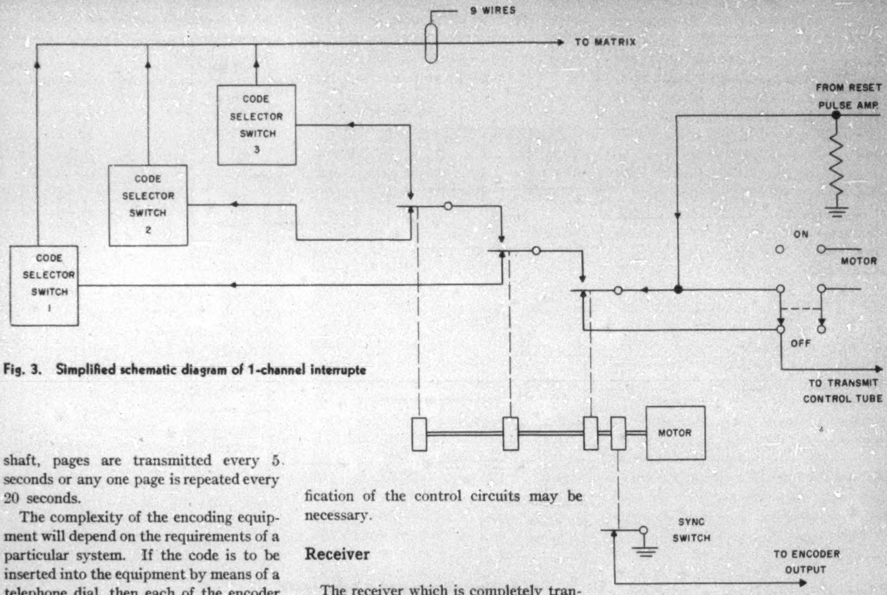


Fig. 3. Simplified schematic diagram of 1-channel interrupte

shaft, pages are transmitted every 5 seconds or any one page is repeated every 20 seconds.

The complexity of the encoding equipment will depend on the requirements of a particular system. If the code is to be inserted into the equipment by means of a telephone dial, then each of the encoder switches will have to be replaced by a register and its associated circuitry. The number of channels required will also make the circuitry more or less complex. In any event, the operations described will have to be performed in one manner or another to generate the required signal.

Transmitter

The ouptut from the encoder is fed over a connecting cable to the transmitter. The transmitter must be capable of being amplitude-modulated with frequencies ranging from 160 to 301 cps. The transmitter must also meet the Federal Communications Commission's requirements, which will vary, depending on the service for which the system is to be licensed.

At present, transmitters with 3, 5, and 30 watts of power input to the final amplifier stage are available for use in the industrial system. The encoder has the necessary facilities for controlling the power to the transmitter and the transmitter carrier. Provision is also made for aurally monitoring the modulation to the transmitter.

Where increased coverage is necessary, transmitters with greater power output are required. Commercial transmitters which are readily available can be used in this application. However, some modification of the control circuits may be necessary.

Receiver

The receiver which is completely transistorized weighs only 7 ounces and is powered by three mercury cells. The receiver case is made of a material having good high-frequency characteristics and has a clip which will hold the unit securely in the pocket or on the belt. Operation of the receiver is limited to a simple OFF-ON switch to apply power and a momentary contact switch to squelch the audio tone when the receiver is signalled.

The transistorized receiver is a 4-transistor unit. The transistors are used in a super-regenerative detector stage, a 2-stage audio amplifier and a "buzzer" or audio-oscillator stage. Also included in the receiver is a decoder which gives each receiver its individuality. The schematic for the receiver is shown in Fig. 4.

Super-regenerative Detector

Used in this stage is a *2N/346* surface barrier transistor. R-f energy is introduced into the receiver via the tank circuit comprised on *L-2* (the antenna coil), *C-12* and *C-13*. Because of the effects of body capacity and the changes in collector capacity from transistor to transistor a large capacitance-to-inductance ratio is used.

R-f oscillations are generated by returning to the emitter a portion of the in-phase voltage developed across the tank circuit. The inductance in the emitter circuit provides a low d-c path to ground for the emitter. The emitter inductance also provides a parallel resonant circuit with *C-11*. This circuit is tuned so that in effect it acts capacitively. Thus, across the antenna circuit there is a capacitive voltage divider, comprised of the effective capacity in the emitter circuit and the collector to emitter capacity of the transistor. The amplitude and phase of the feedback are controlled by the tuning of the emitter circuit.

The transistor is forward-biased by returning the base back to the negative side of the battery through *R-10*. When the transistor begins to oscillate, the positive excursions of the emitter will allow the charge on *C-10* to increase until the transistor is cut off. At this time the oscillations will stop and remain stopped until the capacitor discharges to a point where oscillations can again be sustained. The *R-C* time constant in the base circuit will determine the discharge time.

The point at which oscillations start is also a function of the strength of the r-f signal. The greater the signal strength, the sooner the oscillations will start. Thus for a given signal strength the quench frequency or the frequency at which the oscillations build up and decay will be a function of the time constant in the base circuit. Also, for a given time constant in the base circuit the quench

Vol.
Pt. 1
195

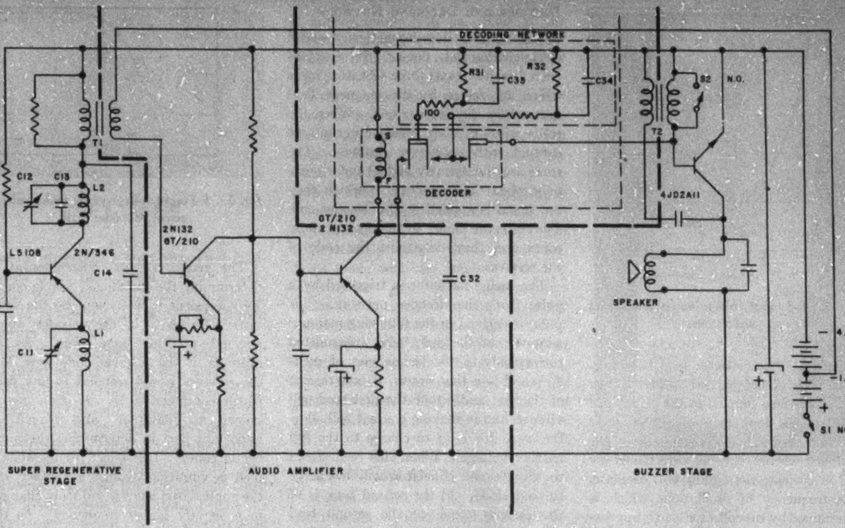


Fig. 4. Receiver schematic

frequency will increase as the signal strength increases.

If attenuation of the audio frequencies is to be avoided, the quench frequency should be greater than ten times the highest audio frequency. Another important factor in selecting the quench frequency is the selectivity. A low quench frequency will result in greater selectivity. The resistor in the base circuit was selected to provide an operating point where the stability and sensitivity were good. The capacitor in the base circuit was therefore selected to provide the proper quench frequency. A quench frequency of approximately 10 kc is employed, which is well above the maximum audio frequency used in the system.

The tuning of the emitter circuit will also have an effect on the quench frequency, sensitivity, and selectivity. Optimum adjustment of the emitter circuit occurs when there is maximum noise output from the detector. After adjusting the receiver for the maximum noise output, the antenna circuit is tuned for the proper frequency of operation. The receivers are pretuned at the factory, thereby eliminating the need for customer tuning.

If the signal varies at an audio rate the collector current will vary at an audio rate. This variation in collector current is detected in the collector circuit of the superregenerative stage. The primary of *T-1* and *C-14* are tuned at the low end of the audio band of frequencies used in the system. The change in collector current develops an audio voltage, across the primary of *T-1*, which is coupled through the transformer to the audio amplifier stage.

Audio Amplifier

The transformer *T-1* matches the collector impedance of the r-f stage to the input impedance of the audio amplifier. Two stages of audio amplification are used, both employing type *2N/132* or *GT/210* p-n-p transistors. Stabilization of the operating point is important and necessary in these stages to prevent an increase in collector current due to temperature variations, and to compensate for variations in transistor parameters between different transistors. Stabilization is accomplished by the use of a battery tap for the base bias of the first stage and by using larger emitter resistors in the emitter circuits of each of the stages. The emitters are bypassed with electrolytic capacitors to prevent degeneration of the audio signal.

The gain of the audio amplifier is controlled by a variable resistor in series with the bypass capacitor in the emitter circuit of the first stage. This method of control allows the audio-frequency gain to be varied without affecting the d-c operating point.

The collector of the first audio stage is direct-coupled to the base of the second stage. The collector load of the second stage is the coil in the decoder. This coil is tuned to the high end of the audio-frequency band by means of capacitor *C-32*. In effect, the tuning of *T-1* and the decoder coil results in a stagger-tuned amplifier which peaks at approximately 256 cps (tone 7).

The operation of the audio amplifier is such that the output is limited on a strong signal. The maximum output is approximately 2.3 volts, rms, on tone 7 with a strong input signal to the antenna. The over-all gain of the audio amplifier is approximately 50 db (decibels).

Audio-Oscillator or Buzzer Stage

The audio-oscillator stage employs an n-p-n type of transistor which is a selected *2N/169*. Referring to the schematic, it is seen that there is no bias applied to the base circuit of this stage. The emitter and the base of the transistor are tied together through the low resistance of the secondary of *T-2*. With no bias, the transistor is practically cutoff. In fact the current is less than 10 μa (micro-

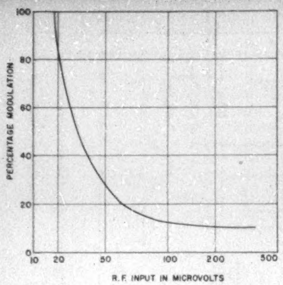


Fig. 5. R-f input voltage versus percentage modulation

amperes). The collector load is a magnetic type of hearing-aid speaker. The collector is also coupled to the primary of *T-2*. If the loop gain of the stage is greater than one and the phase of the feedback to the base as determined by *T-2* is correct, the circuit will oscillate. The frequency of oscillations which is determined by the collector load is approximately 1,500 cps. However, under the stated biasing conditions, the loop gain will be less than one and the oscillator will be blocked. This is the normal operating condition.

If a positive pulse should appear on the base of this stage, it will be amplified and returned in phase to the base via *T-2*. If the pulse is of sufficient amplitude, the instantaneous gain to the transistor will reach such a value that the loop gain will be greater than one. At this time, the transistor will be driven into oscillation and will remain in oscillation. Once the oscillations have started, they may be stopped by depressing a momentary contact switch which is across the secondary of *T-2*. This automatically resets the oscillator and returns it to the stand-by condition to await the next positive pulse. The positive pulse to trigger the oscillator is derived from the decoding network when the correct code has been received.

In the blocked condition this stage will draw less than 10 μa; however, in the oscillating condition the stage will draw up to 6 ma (milliamperes). The current drawn varies from transistor to transistor since the operating point of each of the transistors may vary. To ensure that the oscillator will not be self-triggering at high temperatures, the transistor is selected for low-leakage current. To ensure that the transistor will oscillate properly when it is triggered, it is also selected for the proper beta characteristic.

DECODER AND DECODING NETWORK

The decoder is a hermetically sealed electromechanical device. It contains (in a 3-reed system) three vibrating reeds which are driven by the magnetic field induced by the decoder coil. When the reeds vibrate, they make intermittent contact with stationary contacts. The reeds are of extremely high Q and vibrate only when the frequency to which they are tuned is applied to the decoder coil. The reeds are tuned to the desired audio tones and they determine the code of the receiver.

The audio oscillator is triggered by a pulse from the decoder network. The pulse is derived in the following manner: assume that the code being transmitted corresponds to the frequencies of reeds 1, 2, and 3 in that order. Upon receipt of the first audio tone the first reed will vibrate, and in making contact will allow the capacitor *C-33* to charge to the full battery voltage. When this reed ceases to vibrate the capacitor will discharge through *R-31*. If the second tone is of the correct frequency the second reed will vibrate and hit its contact. When this occurs, the charge on *C-33* will transfer to *C-34*. The capacitor *C-34* will also discharge after the second reed ceases to make contact. If the third tone is the correct one, the third reed on making contact will transfer the remaining charge to the base of the audio oscillator, causing it to oscillate. The individual reeds will vibrate and make contact at any time that their particular frequency is received. However, the tones must be in the proper sequence to transfer a positive pulse to the buzzer stage.

As mentioned earlier, the nine audio frequencies used were determined by the decoder. To eliminate any harmonic problems in connection with the reed response, the nine audio frequencies utilized lie within a single octave. The exact frequencies chosen represent a compromise of several considerations.

As the frequency is increased, reed sensitivity decreases and more driving power is required to achieve a given deflection amplitude. Also, as the frequency decreases the response time of the reed increases and the time required for the reed to reach a given deflection amplitude increases. The natural frequency of a vibrating reed fixed at one end varies inversely as the square of its length; therefore the physical size of the reed had to be considered. The compromise resulted in selecting a group of frequencies ranging from 160 to 301 cps.

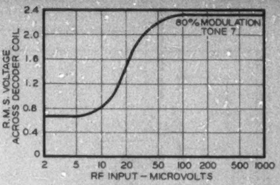


Fig. 6. R-f input voltage versus audio voltage across decoder coil

The reeds are also instrumental in determining the number of usable codes. Two adjacent tones cannot be the same since both reeds in the decoder would respond at the same time. As an example, if the receiver code were 115, the receiver would respond to any code which contained a 15. A phenomenon known as "hang-on" also played an important part in determining the number of codes. This effect occurs when a reed is vibrating against a contact and the applied frequency is suddenly changed to a slightly higher frequency. As the reed vibrates against the contact, its effective length is shortened and therefore its resonant frequency increases. At the same time the Q of the reed decreases. If the applied frequency is suddenly changed to a slightly higher frequency, the reed will continue to vibrate. The end result would be false triggering of the reed. This will be more evident in the following example. If the code 124 is being transmitted, the receiver with code 124 will respond in the normal fashion. At the same time the receiver with the code 214 could respond due to the "hang-on" of the middle reed. To alleviate this condition codes with ad-

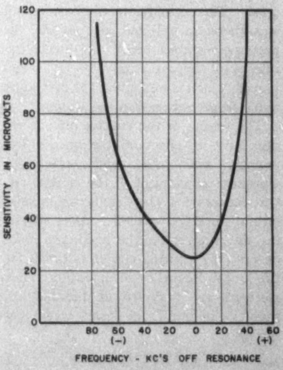


Fig. 7. Selectivity

Vol.
Pt. 1
195

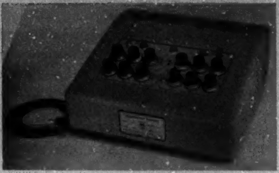


Fig. 8. Encoder

jacent descending numbers are eliminated from the system. These effects limit the number of usable codes in a 3-reed system to 455, and in the 4-reed system to 3,235.

The time intervals of the coded signal as shown in Fig. 1 were also determined by the reeds. The time duration of the tones must be long enough to allow the reeds to make contact. Since the reeds have Q's in the order of 200 they will not make contact immediately. The gaps between the tones are inserted to allow the reeds to decay somewhat before the next tone is applied to the decoder coil. With the gaps the hang-on effect of the reeds is less likely to occur. The guard time is necessary to allow the capacitors in the decoding network to discharge before the next call is received.

Power Supply

The receiver is powered by a 4-volt battery made up of three mercury cells connected in series. A tap on the battery provides the bias for the first audio stage. With the power switch in the OFF position the battery drain is approximately 10 μa. With the power switch on and the buzzer off the battery drain is approximately 750 μa. This is the normal operating current. The current drain with the buzzer on is approximately 6 μa.

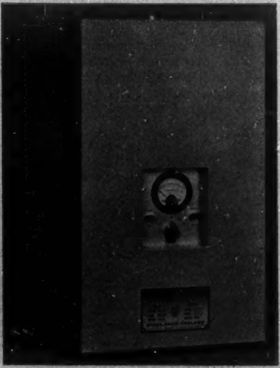


Fig. 9. Transmitter

With a normally operating current drain of less than one ma the battery life is approximately 950 hours. The audio-oscillator stage draws current only when the receiver is triggered by the correct code and, since the triggering is done intermittently, the additional current drawn by the audio oscillator will not seriously effect the battery life. Tests made with the receiver operating continuously and with various duty cycles have resulted in approximately the same operating battery life.

A new battery will have a voltage slightly greater than the 4 volts but, after the first 50 hours of operation, the battery voltage will decrease to 4 volts and remain at 4 for 90% to 95% of its life.

Receiver Variations

Several models of the receiver which differ in the number of reeds in the decoder and the operating frequency are currently in production. Three-reed or 4-reed receivers differ in the decoder and the decoding network constants. Receivers operating at 27 or 35 mc (megacycles) differ in the antenna coil and the constants in the base circuit of the *2N/346*.

Receiver Testing and Performance

The receivers are tested by inserting them in an upright position in the center of a loop. The loop is fed directly from an r-f signal generator which is amplitude-modulated with a signal from a tone-sequence generator. The receivers are required to trigger with a signal of approximately 50 μvolts into the loop. The receivers are also checked with adjacent and reverse codes to ensure that they will not false-trigger in the field.

Typical performance curves of 27-mc 3-reed receivers are shown in Figs. 5 through 7. Fig. 5 shows the required signal into the loop to trigger the receiver for various percentages of modulation. With 100% modulation as the reference or zero-db point, a modulation percentage of only 60% will result in a 2-db loss in receiver sensitivity.

The audio voltage developed across the decoder coil for various amounts of signal into the loop is shown in Fig. 6. This curve shows the limiting action of the receiver for input signals above 50 μvolts.

The selectivity of the receiver is shown in Fig. 7. The average 6-db bandwidth is approximately 75 kc for the 27-mc receivers. The 35-mc receivers have a bandwidth of approximately 45 kc.

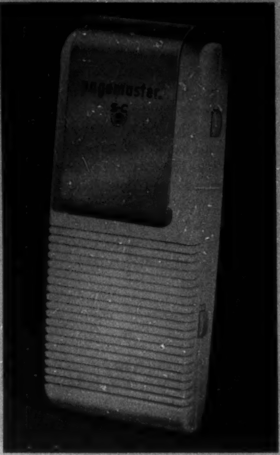


Fig. 10. Receiver

Industrial Equipment

The equipment used in industrial and hospital paging systems is shown in Figs. 8 through 10. The encoder is designed in an attractive case for location near the paging operator. It is 14 inches wide, 14 inches deep, and 7 inches high. The encoder has a self-contained power source, a transmitter remote control switch and a carrier monitor light. There are 4 code group switches and 12 tone-selecting knobs, three for each of the code groups, which can be set up simultaneously. All vacuum-tube circuits with the exception of the transmitter control circuit and the reset pulse amplifier are of the plug-in type. This encoder is capable of generating the codes required to service 455 selective receivers.

The transmitter pictured is the 30-watt version. The 5-watt unit is of the same general construction but somewhat smaller in size. The 5-watt transmitter is $16^{1}/_{4}$ inches high, $8^{1}/_{4}$ inches wide, and 7 inches deep. The transmitters, which are designed for wall mounting, can be remotely located and controlled from the encoder.

The receiver shown in $2^{3}/_{4}$ inches wide, $5^{3}/_{4}$ inches long, and 1 inch deep. The case which is styled for easy insertion into the pocket is ivory-colored with a gold clip for securing the receiver in the pocket or on the belt. A twist-lock cap over the battery compartment allows the battery to be easily replaced.

At present, a majority of the industrial systems in operation are operating at 27.255 mc in the citizens radio band. The range or coverage in any given installation is dependent on such factors as the antenna height and terrain conditions. In buildings the coverage depends on the type, layout, and construction of the building. Each installation has specific requirements and it is difficult to make broad statements as to exact coverage without having made a survey. However in nearly all installations where one building or a group of buildings in the same area are to be covered, the 5-watt transmitter will provide the necessary coverage. Where increased coverage is required the 30-watt unit is available.

Conclusions

The development of the Pagemaster system is a further step toward improving communication and paging operations, and offers many advantages over other types of paging systems.

The system can be easily and economically installed. A system can be started with a few receivers, and expanded at any time as the need arises without any additional changes in the base station equipment. The paging operation is quiet, personal, and selective. The system provides instant contact with key personnel.

Reference

1. City-Wide Personal Signalling at Allentown-Bethlehem, Pa., C. R. Kraus. See pp. 52–55 of this issue.

City-Wide Personal Signaling at Allentown-Bethlehem, Pa.

C. R. KRAUS
MEMBER AIEE

IN THE Allentown-Bethlehem area of Pennsylvania, the strange sound of a low whistle issuing from the coat pocket of a pedestrian (who will, on hearing it, probably seek the nearest telephone and dial a number) might be mystifying to the uninitiated. But to the pedestrian from whose pocket the whistle actually came, it represents a new Bell system telephone service that for some time has been functioning in this area.

Known as Personal Signaling Service, this new development in the telephone art is consistent with one of the basic objectives in furnishing complete telephone service, i.e., making certain that the maximum number of calls are completed and that the calling party reaches the person with whom he wishes to speak.

Need for Signaling Service

To attain such a dual objective successfully obviously suggests the need for reaching the called party when he is not in the immediate vicinity of his telephone. One effort in this direction has resulted in the provision of extension bells which the telephone company has been furnishing for many years. Others have resulted in the provision of code calling or paging systems by means of which, through audible tones or voice, the wanted person may be called to the telephone if he is anywhere on the premises. Such devices have definite limitations, however, and they are useless when the desired individual is away from his premises. Selective signaling service to receivers mounted in vehicles is another useful Bell system communications device directed toward this same objective.

Radio paging pocket-carried receivers which are signaled on a nonselective basis by voice operation also offer possibilities and have a definite field of use since with them the transmission of messages as well is possible. Unfortunately, however, these pocket-carried receivers also have several limitations. First, it is required that the user take certain action at rather frequent intervals; he must turn on his receiver and listen through a cycle of codes or messages to find out whether or not he is being paged. Not only is the individual inconvenienced but also, and of much greater significance, he is unavailable during the period when he is not listening to his receiver. Second, there is a limit to the number of customers who can be served over a single channel. This limit is the natural consequence of the number of times a particular code or message must be repeated in order that the individual being paged will hear it. As the number of customers to be served increases and the number of messages to be put on the air simultaneously becomes greater, the length of the listening cycle increases proportionately.

By contrast, the user of the selective signaling system hears only his own signal when it is sounded over his pocket-carried receiver. No listening cycle is involved; the signal is received at virtually the same time it is transmitted. The service also has an advantage over the vehicular selective signaling service because it eliminates the requirement that the user remain in the proximity of his vehicle. The advantage of selective signaling to pocket-carried receivers, therefore, is to increase substantially the ability of calling parties to communicate with the individual carrying the receiver. As a result, the system is ideally suited to doctors, lawyers, and other professional men, as well as to many types of businesses which must maintain fast, certain, and economical contact with their personnel.

How the Service Functions

A brief, general description of how the service functions may best be understood by reference to Fig. 1. Assuming that the customer is wanted, that he presumably is somewhere in the area, and that he has with him his pocket-carried signaling receiver, the steps in the operation of the service are as follows:

Step 1. By prearrangement, his secretary, or some other person, dials a special number and reaches the telephone company's signaling operator. The secretary advises the operator of the number to be signaled. This is a 4-digit number.

Step 2. The signaling operator "dials" the 4-digit number. She also records the information for billing purposes.

Step 3. Four voice-frequency tones are transmitted by the radio transmitters over the 35-mc (megacycle) channel.

Step 4. The customer's pocket receiver

Paper 58-813, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 31, 1958; made available for printing July 2, 1958.

C. R. Kraus is with The Bell Telephone Company of Pennsylvania, Philadelphia, Pa.

Vol.
Pt. 1
1959

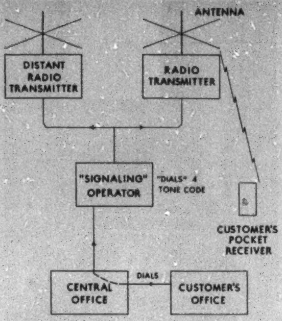


Fig. 1 (left). Block diagram of Personal Signaling Service

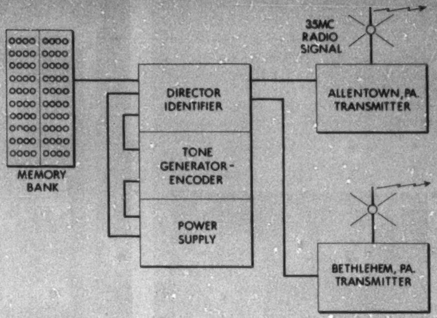


Fig. 2 (right). Block diagram of control equipment

is actuated by the four voice-frequency tones and the customer receives the signal, in the form of a pleasant whistle.

Step 5. On being alerted, the customer presses the cutoff button to stop the whistle and calls, by prearrangement, his office or some other point from the nearest telephone.

Inception

The telephone company offers three types of mobile telephone service. The first of these, General Mobile Service, provides for 2-way telephone communication between a mobile unit within range of a radio base station and any wire telephone or between two such mobile units. The second, Dispatching Service, provides for 2-way telephone conversation between a designated dispatching office of a customer in a mobile service area and a designated mobile unit of the customer within the range of the radio base station. The third, Selective Signaling Service, is a 1-way service from any wire telephone located in a mobile service area to a mobile unit equipped for such service within range of the land radio station serving that area.

Selective Signaling Service is furnished only to receivers mounted in vehicles. The Allentown-Bethlehem Personal Signaling Service is an extension of this same type of service to pocket-carried receivers.

Comparison of Voice, Private, and City-Wide Systems

In any discussion of this new selective type of service, there is the obvious question of how it fits in with existing voice paging services and other types of selective signaling systems which stem from the basic concept of a modulating system for transmitting a number of voice-frequency tones and a receiver which is individually sensitive to a specific succession of certain tones.

The voice system, of course, satisfies a definite need and fulfills requirements so far as the transmission of a message is concerned. However, with a large number of receivers in a system, the voice system can be inadequate because of the large amount of voice traffic. By contrast, inherent in a selective signaling system is the capacity for handling a large number of customers. For example, the system installed in the Allentown-Bethlehem area is a city-wide signaling system with a capacity of approximately 3,200 receivers, any one of which can be signaled. This means that the system capacity is sufficient to make the service available to all prospective users in a sizable service area.

There are many places where a system with a capacity for a smaller number of receivers and with a fewer number of digits to be transmitted is suitable, such as individual buildings, manufacturing plants, hospitals, and similar locations. A description of these simpler private systems, together with information on the fundamental modulating system and the details of the transistorized receiver with its selective mechanism, is covered in a companion paper elsewhere in this issue.[1]

Objectives

The objectives of the experimental installation of personal signaling service in the Allentown-Bethlehem area are in two categories, commercial and technical. Commercially, the area was selected as a good location to explore the possibilities of the field of use of a city-wide system. It was installed to determine the types and the number of customers who would find a use for the service, as well as the amount of use which the customers would make of the service. Fundamentally, the question to be determined is whether this sort of a system will find sufficient customer acceptance to justify its existence.

Technically, the objectives were to determine the quality of performance of the equipment components of the system and to obtain sufficient experience in service to formulate basic operational needs.

Description of the System

The Allentown-Bethlehem system consists of two radio base station transmitters arranged to signal pocket-carried receivers selectively. In addition to the transmitters, control equipment as shown in Fig. 2 is part of the system. A device called a memory bank is installed at the switchboard and the operator selects the voice tones to be transmitted by setting

Fig. 3. Transmitter

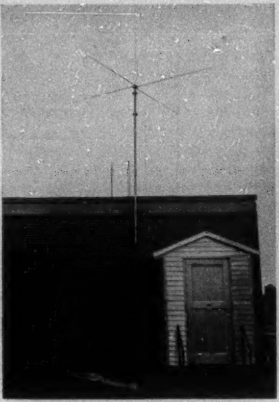


Fig. 4. Transmitter housing and antenna

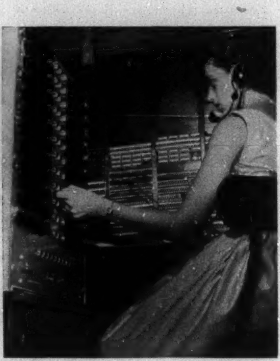


Fig. 5. Memory bank

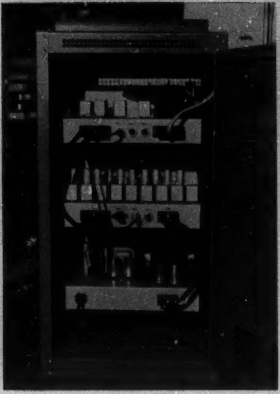


Fig. 6. Tone generator and associated equipment

Top—Director unit
Center—Tone generating unit
Bottom—Power supply

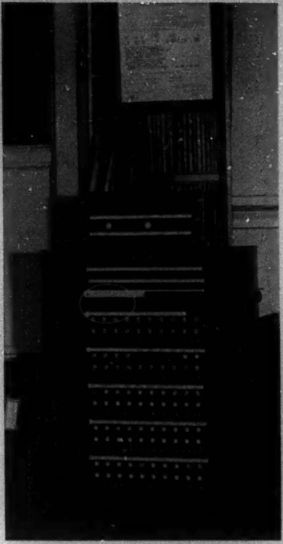


Fig. 7. Other control equipment

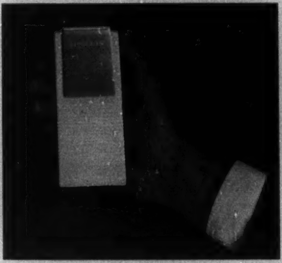


Fig. 8. Receiver

dials on the memory bank. The tone generators and other equipment modulate the radio-frequency carrier and perform other necessary functions.

The amplitude modulation transmitters (see Fig. 3) are Erco Model *301TA*, having an output of about 250 watts and designed to operate in the 30-to-40-mc band. They are operated simultaneously, one at 35.3378 mc and one at 35.3412 mc. This separation of 3.4 kc between transmitter frequencies removes the beat note from the signaling tone band. The antenna structure associated with each transmitter is of standard design consisting of an Andrew type *900* folded unipole antenna. Fig. 4 shows one transmitter housing and antenna on the roof of the Allentown telephone central office building. The other transmitter is located in the Bethlehem telephone central office building.

CONTROL EQUIPMENT

The memory bank, shown in Fig. 5, is part of the control equipment, and consists of a panel containing dials for selecting and storing codes to be transmitted. A maximum of 24 different codes may be selected at one time. The codes selected are stored and each is transmitted three times sequentially by means of a self-contained, electrically operated scanning mechanism. The scanning is continued as long as codes are present in the memory bank. When the scanning is stopped, station identification is transmitted and both transmitters are shut down.

Fig. 6 shows the tone generator and associated equipment, mounted in the equipment room on the second floor of the Allentown central office. The tone-generating unit makes available nine different audio frequencies, quartz-crystal controlled, in the range of 160–300 cycles. The frequencies are selected in groups of four tones for individual calling codes. The four tones comprising a code are transmitted sequentially and require approximately 2 seconds for transmission. With this signaling arrangement, it is possible to select one individual pocket-carried radio receiver out of a group of over 3,200 such receivers.

The director unit, also shown in Fig. 6, controls the radio transmitters, and arranges for the scanning of the codes in the memory bank and for transmission of the codes to the radio transmitters. Station identification by international Morse Code is automatically transmitted over one of the transmitters once every half-hour during continuous transmission. In addition, station identification is automatically transmitted immediately prior to any period during which both transmitters have been shut down. The radio transmitters are remotely controlled by means of a direct voltage applied over wire lines.

Vol.
Pt. 1
195

Other control equipment consists of two keys and two lamps installed in the test center on the fourth floor of the Allentown central office. One key is associated with each transmitter. The keys are normally in the ON position. When they are turned to the OFF position, they prevent the transmitters from being placed on the air. The lamps indicate when either of the transmitters is turned off; see Fig. 7.

RECEIVER

The receiver shown in Fig. 8, is of a transistorized, superregenerative type and weighs about 8 ounces. It includes an electromechanical selective mechanism consisting of four reeds which respond to individual frequencies and a hearing-aid type of speaker for signaling.[1]

COVERAGE AND TRANSMISSION DESIGN

A discussion of the transmission design and a review of the coverage results, together with certain important data on shielding by buildings at both 30-to-40 and 150 mc, will be found in a companion paper.[2]

Operating Experience

As the result of more than 10 months of operation, it is concluded that

1. The service fulfills a need, as evidenced by extensive usage and a steady growth in the number of customers.
2. The maintenance problems, particularly concerning the receiver are somewhat greater than expected, but further work is in progress and improvement is anticipated.
3. The size, shape, and weight of the receiver might be improved.
4. There are some service problems in regard to handling equipment outages.

The experiences thus summarized will be of benefit in the further development of the system.

Contemplated Improvements

A number of improvements in operation are contemplated for the service, some of which are as follows:

1. Dial operation at the switchboard rather than manual setting of the four dials for each call.
2. Use of existing mobile 35-mc systems with signal selective calls interposed between conversations.[2]
3. The use of suitable frequency-modulation personal receivers having a selective signaling arrangement: the use of the frequency-modulation stations now in existence awaits the availability of these receivers.

Conclusion

Personal Signaling Service has established itself in the Allentown-Bethlehem area as an essential part of a telephone service that is ever expanding in scope. From an operational standpoint, the experiment disclosed areas in which improvement is desirable. Some development work remains to be done, therefore, in order that the service may merit a high degree of over-all customer acceptance.

References

1. PAGEMASTER RECEIVER AND MODULATION EQUIPMENT, J. W. Young. See pp. 45–52 of this issue.

2. SYSTEMS ENGINEERING OF PERSONAL RADIO SIGNALING SYSTEMS, Walter Strack. See pp. 55–59.

Systems Engineering of Personal Radio Signaling Systems

WALTER STRACK
NONMEMBER AIEE

IN two other papers Kraus[1] and Young[2] describe in some detail the operational and equipment features of 35-mc (megacycle) amplitude-modulated systems for providing one-way selective signaling service. This paper covers a number of the problems associated with the planning and evaluation of systems operating in either the 35- or 150-mc range which might be used to provide city-wide coverage. Fig. 1 is a simplified schematic of a signaling system and illustrates three of the major areas to be discussed. These areas are the pocket receiver, the radio transmission path, and finally the central office facilities.

Evaluation of Pocket Receivers

In order to plan a radio system it is necessary to know the characteristics of the various components. Many of the equipments used in providing one-way signaling service are fairly conventional and their characteristics are either readily available or can easily be determined. Some new problems exist, however, with regard to the pocket-sized radio receivers. First, the antenna in these units is contained within the receiver case and cannot readily be disconnected to facilitate many of the usual measurements. In the case of the Stromberg-Carlson receiver, the antenna is actually a portion of the tuned circuit in the superregenerative detector and cannot be disconnected without disrupting the operation of the set. The second consideration is that these receivers will normally be operated in close proximity to the user's body. This may affect the performance of the unit by modifying the apparent antenna characteristics or shifting the tuning of the receiver. For these reasons, the sensitivity must be determined, initially at least, when the receiver is carried by a person.

In making tests of this type a signal generator was used as the r-f (radio-frequency) source and the audio signaling frequencies were provided by a special "test-tone generator." The output of the signal generator was fed to an outdoor antenna and the output was adjusted to the level required to just trigger the receiver as the user assumed various orientations at a point about 100 feet removed from the antenna. The actual field strength for various settings of the signal generator were determined using conventional measuring equipment.

Tests have shown that sensitivity is a function of where the receiver is carried, on the belt or in a pocket, and also depends to some extent on the type of clothing worn. Sensitivity variations in the order of ±10 db (decibels) have been measured as the user rotated through 360 degrees.

The procedure for determining the sensitivity of the receivers which have speech output rather than selective signaling is somewhat more involved. Historically, the sensitivity of voice equipment is usually measured in terms

---

Paper 58-989, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted July 31, 1958; made available for printing September 5, 1958.

WALTER STRACK is with Bell Telephone Laboratories, Inc., New York, N. Y.

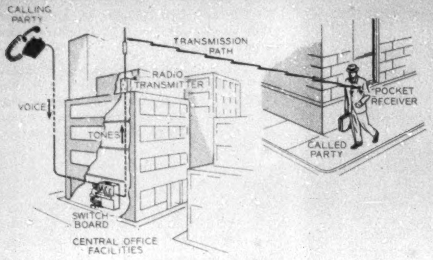


Fig. 1. Personal radio signaling system

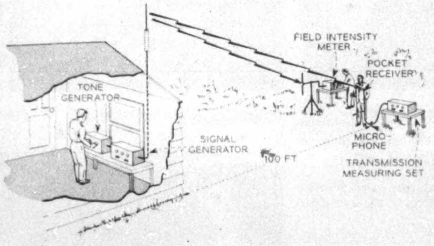


Fig. 2. Procedure for determining pocket receiver sensitivity

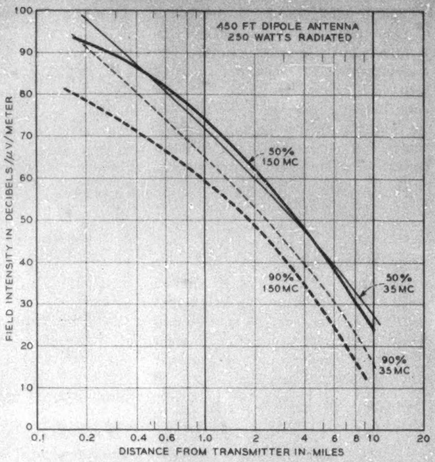


Fig. 3. Field intensity in city streets, 35 and 150 mc

of noise quieting or the r-f level required to produce a given signal-to-noise ratio. The usual arrangement is to bridge a transmission measuring set across the audio output terminals. When this approach was attempted with the pocket receivers, it was found that the meter leads acted as antennas and modified the sensitivity of the receivers in a more or less random fashion depending on the drape of these leads.

One solution to this problem is to use acoustic coupling between the receiver and the measuring set. The output of the miniature receiver speaker can be fed through a length of plastic tubing to a microphone. The output of the microphone is connected to the transmission measuring set thus providing means for measuring either noise quieting or tone-to-noise ratios. This arrangement is illustrated in Fig. 2.

These techniques were found to be satisfactory for original laboratory investigation of receivers of this type. They are not satisfactory, however, from the standpoint of the technician who must maintain a large number of the units on a routine basis. To facilitate this type of work, jigs have been devised to mount the receiver in a fixed position. The r-f signal is introduced using either inductive or capacitive coupling from a near-by lop, probe, or the like. This whole arrangement can be placed in a shielded box or room to exclude extraneous signals and can be calibrated to give sensitivity data of the type available when the more elaborate out-of-door setup is employed.

Transmission Considerations

To ensure satisfactory transmission performance of a radio system the following simple relationship must exist: transmitter output power−path loss=field intensity≥receiver sensitivity.

Essentially, this states that whenever the field intensity is equal to or greater than the sensitivity of the receiver, a satisfactory grade of service will be provided. In order to determine the useful range of the system, one is concerned with the relationship between field intensity and the physical separation between the transmitter and the receiver.

A considerable quantity of data on path loss has been accumulated over the last few years as a result of the operation of mobile telephone services at 35 and 150 mc. This information is restricted, however, to those areas into which an automobile might travel. Additional data are required to take into account the loss which might be encountered by a subscriber to a personal signaling system when he enters a shielded area such as a building.

This new information has been obtained in such a way that performance within buildings can be predicted by adding a factor, the building loss factor, to the earlier data on field intensity in city streets.

Field Intensity in City Streets

Fig. 3 illustrates some of the data on field intensity in city streets at 35 and 150 mc which has been obtained in the New York City area.[3]

Because of the wide fluctuations which exist as a result of the many obstructions in the path, it was necessary to consider not only the median values (curves labeled 50%) but also the distribution of the radio field strength. The lower curves in Fig. 3 represent the minimum field intensity that can be expected in 90% of the cases at a given distance from the transmitter. The curves shown represent many hundreds of samples of signal strength. It is of interest to note that the median field intensity at 35 and 150 mc is nearly equal. However, variations in level in city streets are somewhat greater at the higher frequency as indicated by the spread between the 50% and 90% curves.

Building Loss Factor

It was anticipated that the amount of loss introduced by a building would be

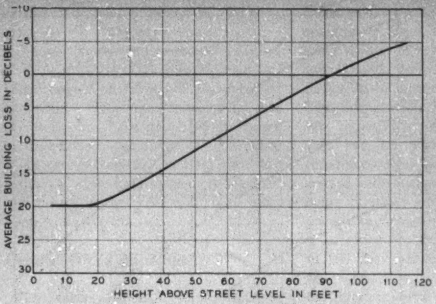


Fig. 4. Building loss versus height above street level

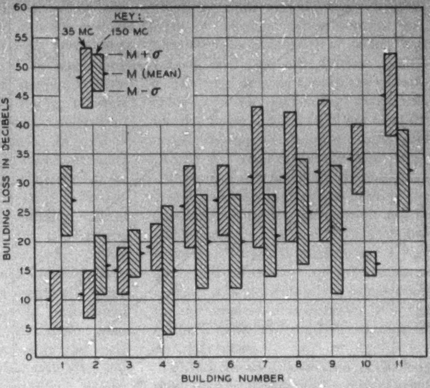


Fig. 5 (right). Measured loss in various buildings in a large city

a function of the type of construction. For this reason, tests were made in a number of buildings, total of 11, having obviously different types of architecture. A number of these had large areas of glass on the lower floors while others used older construction techniques and had fairly thick walls of brick, concrete, etc., and employed structural steel framework.

Preliminary testing in a few buildings showed that the weakest signals, and therefore the greatest path losses, were experienced on the street-level floors of buildings (excluding the basements). An example of the variation in loss on various floors of a building is shown in Fig. 4.

Any system then which will provide adequate service on the street-level floor of a building can thus be expected to be completely adequate on upper floors. Qualitatively, this seems reasonable since the shadowing effect of surrounding buildings decreases as one moves up in a building.

For this reason the bulk of the data on building loss was taken on the street-level floors. Portable field intensity equipment which had adequate selectivity to reject signals on adjacent frequencies was employed to measure the field intensity produced by standard 250-watt mobile telephone transmitters. The number of measurements taken varied from building to building depending on the amount of floor space and the complexity of the floor plan. Measurements were also made in the streets adjacent to each of the buildings. The difference between the median field intensity in the street and the field intensity on the main floor has been called the building loss.

In general, the variations in field strength within a building tended to follow a logarithmic normal distribution. It was possible, therefore, to express building loss in terms of a median value and the standard deviation.

A summary of the results for the 11 buildings is shown in Fig. 5.

It was found that in certain cases a qualitative prediction might be made concerning the losses based on the architectural characteristics of a building. For example, buildings 2, 3, and 4, which have losses well below the average, had high ceilings and appreciable areas of glass. Building 11 which had low ceilings and small windows has a rather high loss.

Unfortunately, such guesses do not always hold true. Building 8 exhibited high loss at both 35 and 150 mc in spite of the fact that it had high ceilings and large window areas.

In order to make use of these data to plan a system, it is necessary to develop a method for describing a group of buildings rather than an individual one. To accomplish this a statistical method was used for combining the characteristics of the 11 buildings so as to obtain a composite picture of the loss introduced by the buildings. This information is shown in Fig. 6. It indicates that the average building loss for this environment is 22 db at 150 mc and 24 db at 35 mc. On the average then, the building losses at 150 mc are slightly less than at 35 mc and as indicated in the figure the variations in loss are also somewhat smaller at the higher frequency. Referring back to Fig. 3, these variations within buildings are just the opposite of those experienced in streets where the low frequency exhibited smaller variations.

Service Range

With information on building loss available, it is possible to estimate the service range possible for systems of the type. Figs. 7 and 8 illustrate the ranges possible at each frequency for varying degrees of service reliability and receiver sensitivity.

For reliabilities in the order of 90%

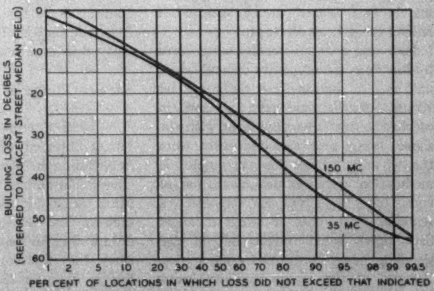


Fig. 6. Variations in building loss, 35 and 150 mc

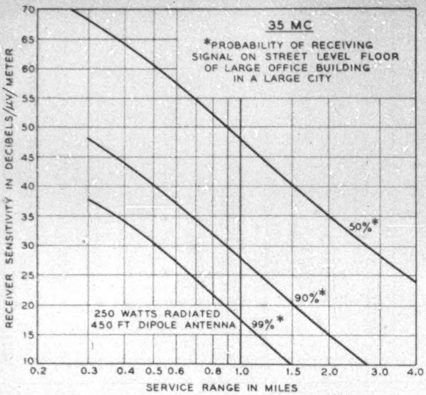


Fig. 7. Estimated service range in a large city, 35 mc

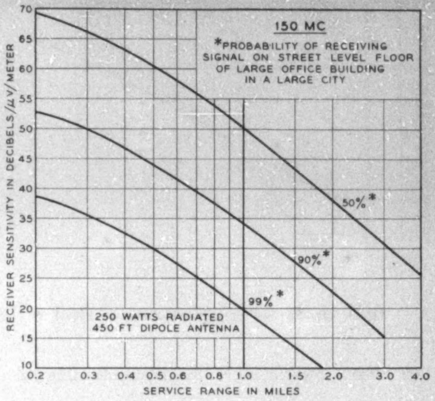


Fig. 8. Estimated service range in a large city, 150 mc

and receiver sensitivities of 100 microvolts per meter (40 db above 1 microvolt per meter), ranges of 0.50 and 0.70 mile are possible at 35 and 150 mc respectively for customers located on the street-level floors of the buildings. On upper floors or in the streets the coverage increases considerably. Based on the data shown in Fig. 3, ranges in the order of 4 miles at 35 mc and 3 miles at 150 mc can be expected in large cities.

Because of the wide variations in loss throughout any one building and because the user may be moving more or less randomly from high- to low-loss areas, a considerable improvement in service can be achieved if the signaling call is transmitted more than once. For example, if the probability of receiving one transmission is 90%, the probability of receiving at least one out of three transmissions is 99.9%. Similarly, the improvement in performance can be stated in terms of an increase in range. By sending the signal three times instead of just once, an increase in range of approximately two to one can be achieved if reliability is held constant.

Increased coverage can also be achieved by installing a multiplicity of land transmitters. As indicated in the paper by Kraus, two transmitters are now in use in the Allentown, Pa., area with the signal being sent simultaneously from the two units. Since this is an amplitude modulation system which makes use of audio tones to accomplish the selective signaling, care must be taken to avoid beat notes which might fall in the signaling band from 160 to 300 cycles. Because of the equipment complexity required to keep the beat note below 160 cycles, the decision in this case was to separate the two transmitter frequencies by approximately 2 kc thus assuring that the resultant beat note would always be above the signaling band. The selectivity of the receivers is such that there is no impairment in receiver sensitivity as a result of this 2-kc separation.

It is also necessary to control the phase of the modulation of the transmitters to prevent cancellation of the audio tones in the receiver. Practically, the author has found that if the difference in phase is held within 45 degrees, there is no decrease in receiver sensitivity when two radio signals of equal intensity are received.

Because of the more stringent frequency tolerances involved at 150 mc, it is doubtful whether similar techniques could be employed in this frequency band. The alternative is to maintain better frequency stability or to resort to sequential modulation of the base station transmitters: first sending the code from one and then repeating it slightly delayed in time from the other.

Service Range at Allentown

It is of interest to note the actual ranges obtained with the systems now in operation in Allentown and Columbus, Ohio. The bulk of the data was obtained using the pocket receivers themselves to indicate the service limits. The procedure was to have trained personnel carry the receivers in various areas a known distance from the land transmitters. The land station was arranged to signal these receivers repeatedly at a rate of two or three times per minute. The service reliability in a given area was then determined by comparing the number of calls received with the number actually transmitted during a 15- to 30-minute interval. The sensitivity of the receivers was checked at frequent intervals and a number of spot checks of field strength were made using conventional measuring equipment.

For reliabilities ranging from 90% to 100%, it was determined that the range of the Allentown system when used outdoors varied from 3 to 5 miles. This was obtained by the use of receivers which had a sensitivity in the order of 40 to 46 db above 1 microvolt per meter.

Experience concerning coverage within buildings is somewhat limited because most of the larger buildings in the Allentown area are located rather close to the transmitter sites. Based on a small sample, the range is fairly close to that which might be predicted, varying from 0.5 to 1.0 mile on the first floors of the buildings investigated.

Commercial service was initiated in Columbus on May 1, 1958. In this case the transmitting antenna location is well above surrounding buildings and this has resulted in extended coverage. The service range to customers in streets is in the order of 12 miles. Data on coverage in buildings are again limited because most of the larger buildings are grouped near the transmitter in the downtown district. Ranges in the order of 5 to 6 miles seem to be typical for the

Vol.
Pt. 1
1959

smaller industrial buildings in the suburbs.

When the pocket receivers are used in an automobile, the useful range depends largely on the exact placement of the unit. With the receiver in the driver's pocket, the useful service range in Columbus varies from 3.5 to 6 miles. If the receiver is fastened up near the sun visor, some improvement is obtained and ranges in the order of 7 to 10 miles seem to be typical.

Central Office Facilities

The earlier papers have described in some detail the type of central office equipment now being used in Allentown, Pa. This installation, by way of review, is a completely manual system. The calling party must contact an operator and provide the correct 4-digit code. This code is placed into the system by the operator who sets four rotary switches and operates a start switch. This information is then translated into the series of four audio tones which are used to modulate the radio transmitters. After the call has been transmitted three times (since other calls are interleaved this may take approximately 3 minutes), the operator removes the call from the system by throwing a switch to the off position. In this system the operator is required to prepare a ticket indicating the numbers of the parties involved and the time of day at which the call was placed.

In order to reduce the operator's work load, the system now in operation in Columbus has facilities to allow the operator to place calls into the system using a dial or keyset from a standard switchboard position. The digits are stored in a series of stepping relays. Two or more of these storage registers can be provided so that a number of calls can be placed into the system simultaneously. Since personal signaling service and mobile telephone service are being provided on the same radio frequency in Columbus, control equipment is provided to permit the two services to use the channel alternately. When the channel is available for signaling, the information in the registers is translated into the sequence of tones which are then directed to the radio transmitters. The information in the registers is automatically erased after the message has been transmitted the required number of times and a lamp then indicates to the operator that the register is available for the next call. As in the Allentown case, the operator must prepare a ticket to allow the call to be charged properly.

Future Development

With the solving of these receiver, transmission, and central office problems, another plateau has been reached and more work lies ahead.

The next logical step, and one that is now under consideration, is that of allowing the calling parties to dial calls directly into the signaling system. From a systems planning standpoint this involves arriving at an arrangement which can work with the many different types of telephone offices now in use. This includes step-by-step, panel, and various types of crossbar systems. Effort is also being directed toward the streamlining of the accounting procedure. Ideally, the goal is to make use of automatic message accounting techniques (AMA) to record the traffic data.

From a transmission standpoint, there is still much to be learned regarding the variations in field strength both in the bands now in use and at higher frequencies. A related problem, and a very important one, is the design of miniature antenna arrangements which can abstract the optimum amount of power from the complex radio field patterns which exist in urban areas.

Finally, there is the problem of devising still smaller receivers: receivers that have a high order of sensitivity, selectivity, and stability, and that are capable of high signaling speeds. These are the challenges before the research, planning, and development groups—technology of the future must be coupled with experience from the past to arrive at solutions which are both operationally and economically satisfactory.

References

1. City-Wide Personal Signaling at Allentown-Bethlehem, Pa., C. R. Kraus. See pp. 52–55 of this issue.

2. Pagemaster Receiver and Modulation Equipment, J. W. Young. See pp. 45–52 of this issue.

3. Comparison of Mobile Radio Transmission at 150, 450, 900, and 3700 MC, W. Rae Young, Jr. *Bell System Technical Journal*, New York, N. Y., vol. 31, Nov. 1952, pp. 1068–85.

Design of Random-Noise Transformers

M. P. VORE
MEMBER AIEE

Synopsis: This paper develops a method of choosing the core required for transformers which must handle random-noise voltages. Experimental verification of this method is briefly described. The early portion of the paper describes briefly some of the characteristics of random noise and outlines the method of calculating the random current produced in a complex impedance by a random voltage.

RANDOM-NOISE voltages and currents are coming into use in a variety of practical ways. This has caused the development of apparatus to generate, amplify, measure, and utilize random-noise power.

One very important recent field of application is in the testing of air-borne military components. The specifications for many of these require testing under conditions of random vibration. The power capability of the electric equipment which supplies random noise power to vibration shakers ranges up to 70 kva.

The design of random-noise power supplies requires a knowledge of the behavior of random voltages, currents, magnetic fluxes, and the relationships among them analogous to that which exists for sinusoidally varying quantities. The design of transformers carrying random-noise power is of particular interest because these components are usually specially designed for the application, have a relatively long delivery cycle, and represent an appreciable proportion of the total cost of the apparatus. All in all, the equipment designer just cannot afford to be wrong in his transformer design and application.

When this author was faced with the necessity of designing a transformer to

---

Paper 57-1015, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted July 1, 1957; made available for printing August 12, 1957.

M. P. Vore is with the Westinghouse Electric Corporation, Baltimore, Md.

handle high-power random noise, a search of the literature provided nothing that was directly applicable to the problem although the fundamental mathematical tools for solving it are well documented.[1,2] It is felt therefore that this paper describing the application of basic theory to the design of random noise transformers may be helpful.

It is helpful to begin by describing some of the basic calculations involved in handling random electrical quantities. Rigorous mathematical justification for the steps taken may be found by consulting the mathematical literature previously cited.

Elements of Random Noise

If a voltage is varying completely at random, neither its value nor its rate of change can be predicted for any given instant.

Similarly there are random-noise currents. The "shot effect" current in an electron tube is a classic example of a random-noise current. The noise voltage generated in a resistor by the random motion of the free electrons in it is an example of a random-noise voltage.

EXAMPLES OF RANDOM NOISE

On an oscilloscope random noise looks like the familiar "hash" or "grass." When applied to a transducer random noise sounds like the hiss of a sensitive radio receiver tuned between stations.

POWER SPECTRUM

A random-noise voltage may be thought of as composed of infinitely many sinusoidal components of all of the frequencies in the noise bandwidth. These components have random phases.

The power content of the voltage may be distributed uniformly or nonuniformly in the total bandwidth of the noise. This distribution with frequency of the noise voltage is known as its "power spectrum." The power spectrum of a random voltage or current forms the starting point for many calculations involving random quantities.

The power spectrum of a random voltage is the power per unit bandwidth that would be dissipated by a 1-ohm resistance to which the random voltage is applied. A similar definition applies to the power spectrum of a random current. In general these are functions of frequency.

In the following, the power spectra of random voltages and currents will be denoted by E_f^2 and I_f^2 respectively, and their rms values by E_r and I_r.

$$E_r^2=\int_0^\infty E_f^2 df \tag{1}$$

$$I_r^2=\int_0^\infty I_f^2 df \tag{2}$$

If the voltage or current has its power spectrum confined within the bandwidth f_1 to f_2 the limits of integration become f_1 to f_2.

Conceivably, the power spectrum of a random voltage could be determined experimentally by connecting a thermocouple voltmeter to a random-voltage source through a filter with a 1-cycle passband. Then, as this 1-cycle passband is scanned through the frequency range, the power spectrum is the average value of the square of each meter reading plotted against its corresponding midband frequency.

It will be apparent by this time that any quantity such as magnetic flux which varies at random has the characteristics of random noise. It is possible and convenient to define a power spectrum for them in an analogous manner.

WHITE NOISE

If a random voltage or current has a constant power spectrum over its bandwidth it is termed "white noise" by analogy to white light which has a uniform energy distribution.

RELATIONS BETWEEN RANDOM VOLTAGE AND CURRENT

When a random voltage having a power spectrum E_f^2 is applied to a frequency-dependent impedance Z, the power spectrum of the resulting current is

$$I_f^2=\frac{E_f^2}{|Z(f)|^2} \tag{3}$$

The rms current is found from

$$I_r^2=\int_0^\infty \frac{E_f^2}{|Z(f)|^2} df \tag{4}$$

For example, suppose a white noise voltage of rms value E_r with a bandwidth from f_1 to f_2 is applied to a series circuit of inductance L and resistance R. The power spectrum of the white noise voltage is

$$E_f^2=\frac{E_r^2}{f_2-f_1} \text{ for } f_1<f<f_2 \tag{5}$$

$$E_f^2=0 \qquad \begin{matrix} f<f_1 \\ f>f_2 \end{matrix} \tag{6}$$

Substituting equation 5 in equation 3, the power spectrum of the current is

$$I_f^2=\left(\frac{E_r^2}{f_2-f_1}\right)\left(\frac{1}{R^2+4\pi^2f^2L^2}\right) \tag{7}$$

It is interesting to note that due to the fact that the impedance is frequency-dependent, the current caused by a white noise voltage is not a white noise current. The mean square value of the current is

$$I_r^2=\int_{f_1}^{f_2} \frac{E_r^2}{f_2-f_1}\,\frac{1}{R^2+4\pi^2f^2L^2}\,df \tag{8}$$

In this case the limits of integration are from f_1 to f_2 only because the power spectrum of the voltage is zero outside of this frequency band.

Carrying out the indicated operations, the rms value of the current is as follows:

$$I_r=\frac{E_r}{\sqrt{2\pi LR}\,\sqrt{f_2-f_1}\Big/\sqrt{\tan^{-1}2\pi\frac{L}{R}f_2-\tan^{-1}2\pi\frac{L}{R}f_1}} \tag{9}$$

The expression in the denominator may be thought of as a "random noise impedance" Z_n.

The power dissipated in the resistor is

$$P=I_r^2R \tag{10}$$

PROBABILITY DENSITY

The probability that a random voltage at any instant has a value between v and $v+dv$ can be defined and written mathematically. It is

$$P=G(v)dv \tag{11}$$

$$P=\frac{1}{\sqrt{2\pi}v_r}e^{-\frac{1}{2}\frac{v}{v_r^2}}dv \tag{12}$$

where

v = the instantaneous voltage
v_r = the rms value of the voltage

The function

$$G(v)=\frac{1}{\sqrt{2\pi}v_r}e^{-\frac{1}{2}\frac{v^2}{v_r^2}} \tag{13}$$

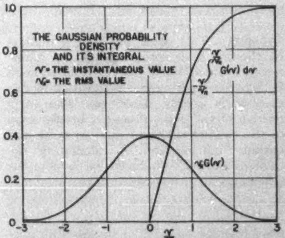


Fig. 1. Plot of the Gaussian probability density and its integral

is the Gaussian probability density function.

The probability that the random voltage is between two discreet values v_1 and v_2 is

$$\int_{v_1}^{v_2} G(v)dv \tag{14}$$

It may be shown that

$$\int_{-\infty}^{\infty} G(v)dv=1 \tag{15}$$

$G(v)$ is called the probability density.

Fig. 1 is a plot of $v_r G(v)$ and of the integral of $G(v)$ between the limits $-v/v_r$ and $+v/v_r$. This latter is the probability that the random voltage at any instant is numerically less than the abscissa times its rms value.

Values of the probability integral

$$\frac{2}{\sqrt{\pi}}\int_0^x e^{-x^2}\,dx$$

may be found tabulated to several decimal places.[3]

Random-noise currents, or magnetic fluxes, also have a Gaussian probability density. That is, for a random current

$$G(I)=\frac{1}{\sqrt{2\pi}I_r}e^{-\frac{1}{2}\frac{I^2}{I_r^2}} \tag{16}$$

or for a random flux

$$G(\phi)=\frac{1}{\sqrt{2\pi}\phi_r}e^{-\frac{1}{2}\frac{\phi^2}{\phi_r^2}} \tag{17}$$

It can be shown that the probability density function is not influenced by the power spectrum of the random quantity. In particular, for instance, a random voltage will produce a random current through an impedance or a random flux in a transformer core, and the resulting current or flux will have a Gaussian probability density.

PRACTICAL PEAK VALUE OF A RANDOM VOLTAGE OR CURRENT

In designing apparatus to handle random noise one must make an engineering decision concerning the maximum peak values of voltage or current or magnetic flux the apparatus will handle. The Gaussian probability density function indicates that there is a mathematical probability that the voltage may be of any size, even infinitely large. The probability that a random quantity will exceed three times its rms value is only 0.3%, however. Therefore, in designing apparatus containing vacuum tubes or transformers or the like having a limited linear operating range one designs so that the linear portion of the range will extend to three times the rms value of the random quantity to be handled.

Transformer Design

CORE FLUX

In designing a transformer for random voltages the designer is faced with the problem of choosing a core that will not saturate on the random voltage peaks of three times the rms value. He could proceed to design a transformer that would handle a sine wave voltage whose crest value is equal to the three times the rms value of the noise voltage at the lowest frequency in the noise bandwidth. Such a transformer would perform satisfactorily but it would be unnecessarily large and uneconomical as will be shown.

When a random voltage is applied to a transformer winding, a random flux is produced in the core. This flux has a Gaussian distribution, and its mean square value may be found by a method analogous to the one just used to find the mean square value of current through an impedance.

$$\phi_r^2=\int_{f_1}^{f_1}\frac{E_f^2}{4\pi^2n^2f^2}df\times 10^{16} \tag{18}$$

where n is the number of turns of the winding.

For instance, if a white noise voltage of rms value E_r and bandwidth from f_1 to f_2 is applied to a transformer winding of n turns, the mean square flux produced in the core is

$$\phi_r^2=\int_{f_1}^{f_2}\frac{E_r^2}{f_2-f_1}\frac{1}{4\pi^2n^2f^2}df\times 10^{16} \tag{19}$$

Working this out, the rms value of the random flux is

$$\phi_r=\frac{1}{2\pi n}\frac{E_r}{\sqrt{f_1f_2}}\times 10^8 \tag{20}$$

For comparison, note that the rms value of the sinusoidal flux produced by a sine wave voltage of rms value E_{rs} and frequency f_s is

$$\phi_{rs}=\frac{1}{2\pi n}\frac{E_{rs}}{f_s}\times 10^8 \tag{21}$$

If one were designing a transformer for a sinusoidal voltage he would pick a crest value of flux ϕ_{ms} (usually limited by saturation) at which he wishes the core to operate.

$$\phi_{ms}=\sqrt{2}\,\phi_{rs} \tag{22}$$

It is convenient to work with flux linkages rather than flux and so the following is written:

$$n\phi_{ms}\times 10^{-8}=\frac{\sqrt{2}}{2\pi}\frac{E_s}{f_s} \tag{23}$$

EQUIVALENT SINUSOIDAL VOLTAGE

White Noise

In the random-noise case the peak value of flux for which one designs is three times the rms value and one designs the transformer so that this is numerically equal to the value considered allowable for the sinusoidal case, i.e.,

$$\phi_{ms}=3\phi_r \tag{24}$$

Thus, the following equivalences result:

$$n\phi_{ms}\times 10^{-8}=\frac{\sqrt{2}}{2\pi}\frac{E_s}{f_s} \tag{25}$$

for the sinusoidal case, and

$$n\phi_{ms}\times 10^{-8}=\frac{3}{2\pi}\frac{E_r}{\sqrt{f_1f_2}} \tag{26}$$

for the white noise case. From these

$$\frac{E_s}{f_s}=\frac{3}{\sqrt{2}}\frac{E_r}{\sqrt{f_1f_2}} \tag{27}$$

Therefore, in practice, to choose a core size for a transformer to handle a random-noise voltage of rms value E_r and bandwidth f_1 to f_2, one defines an equivalent sinusoidal voltage E_s and frequency f_s as in the foregoing and chooses a core size that would be satisfactory for that.

It can be seen that this analysis allows the use of a considerably smaller core than would have been chosen on the basis of capability to handle a sinusoidal voltage of crest value equal to the peak value of the noise voltage and of frequency equal to the lowest frequency in the noise bandwidth as follows.

Core size is proportional to the flux linkages it handles provided the turns remain constant. Hence if instead of choosing a core size proportional to

$$n\phi_{ms}\times 10^{-8}=\frac{3}{2\pi}\frac{E_r}{\sqrt{f_1f_2}} \tag{28}$$

it had been chosen proportional to

$$n\phi'_{ms}\times 10^{-8}=\frac{3}{2\pi}\frac{E_r}{f_1} \tag{29}$$

the core would have been unnecessarily large by a factor of $\sqrt{f_2/f_1}$. Evaluating this for a bandwidth even as narrow as 30 to 300 cycles, the ratio is $\sqrt{10}=3.16$, which is considerable.

The reason for this is that the flux in the core is proportional to the area under a positive or negative voltage loop. The probability that such a loop will have a crest value equal to three times the rms value and simultaneously have a width corresponding to the lowest frequency in the noise bandwidth is considerably less than the probability that the voltage will simply reach a crest value of three times rms.

Shaped-Noise Spectrum

A slightly more complicated case that has been worked out is the calculation of flux caused by a white noise voltage of bandwidth f_1 to f_2 that was first put through a network having a frequency response decreasing at the rate of 6 decibels per octave with increasing frequency.

In this case the frequency spectrum of the voltage is

$$\frac{E}{E_0}=\frac{1}{f} \tag{30}$$

where E_0 = a constant.

Its mean square value E_r^2 and power spectrum E_f^2 are related as follows:

$$E_r^2=\frac{1}{f_2-f_1}\int_{f_1}^{f_2}\frac{E_0^2}{f^2}\,df=\int_{f_1}^{f_2}E_f^2df \tag{31}$$

From this it can be seen that the power spectrum of the voltage is

$$E_f^2=\frac{1}{f_2-f_1}\;\frac{E_0^2}{f^2} \tag{32}$$

One now evaluates E_0 in terms of the observable rms value E_r as follows:

$$E_r^2=\int_{f_1}^{f_2}E_f^2df \tag{33}$$

$$E_r^2=\int_{f_1}^{f_2}\frac{1}{f_2-f_1}\;\frac{E_0^2}{f^2}\,df \tag{34}$$

$$E_r^2=\frac{E_0^2}{f_2f_1} \tag{35}$$

Substituting the value of E_0^2 in the expression for E_f^2 (equation 32) gives

$$E_f^2=\frac{f_2f_1}{f_2-f_1}\;\frac{E_r^2}{f^2} \tag{36}$$

which is more practical to work with because one can observe E_r with an rms reading meter.

Now it is possible to compute the mean square core flux by the method used earlier.

$$\phi_r^2=\int_{f_1}^{f_2}\frac{E_f^2}{4\pi^2f^2n^2}\,df\times 10^{16} \tag{37}$$

$$\phi_r^2=\int_{f_1}^{f_2}\frac{f_1f_2}{f_2-f_1}\;\frac{E_r^2}{4\pi^2n^2}\;\frac{df}{f^4}\times 10^{-16} \tag{38}$$

Working this out gives

$$10^{-8}n\phi_r=\frac{1}{2\pi\sqrt{3}}\;\frac{E_r}{\sqrt{f_1f_2}}\sqrt{\frac{f_2}{f_1}+1+\frac{f_1}{f_2}} \tag{39}$$

For comparison note the expression for rms flux linkages for the white noise case

$$10^{-8}n\phi_r=\frac{1}{2\pi}\;\frac{E_r}{\sqrt{f_1f_2}} \tag{40}$$

and it is seen that in all cases the flux caused by this shaped-voltage distribution is greater than it is for the white noise case. This is as it should be because in this latter case the voltage spectrum is weighted toward the low-frequency end.

By proceeding as before, it is possible to find an equivalent sine wave voltage and frequency that will produce a crest value of flux linkages equal to three times the rms value of noise flux linkages. This is used as the basis of choice of core size.

Utilizing previous work,

$$\begin{aligned} n\phi_{ms}\times 10^{-8}&=\frac{\sqrt{2}}{2\pi}\;\frac{E_s}{f_s}\\ &=\frac{3}{2\pi\sqrt{3}}\;\frac{E_r}{\sqrt{f_1f_2}}\sqrt{\frac{f_2}{f_1}+1+\frac{f_1}{f_2}} \end{aligned} \tag{41}$$

This simplifies to

$$\frac{E_s}{f_s}=\frac{3}{\sqrt{6}}\;\frac{E_r}{\sqrt{f_1f_2}}\sqrt{\frac{f_2}{f_1}+1+\frac{f_1}{f_2}} \tag{42}$$

Experimental Verification

The validity of this method has been experimentally verified in both the case of white noise and noise with a sloping power spectrum. The method used was to establish for an existing transformer the value of 60-cycle sine wave voltage at which the core entered saturation. This was done by observation of magnetizing current. Next, by observation of magnetizing current the value of noise voltage at which the core entered saturation was observed. The results were in good agreement with equations 27 and 42 respectively. It was quite easy to tell when saturation occurred in the random-noise case because occasional peaks would cause violent deflections of the magnetizing current ammeter when the rms noise voltage was too great.

TURNS RATIO

So far only the problem of choice of core size in random noise transformers has been dealt with. Turns ratio, of course, is the ratio of rms voltages just as in the sinusoidal case.

COPPER AND CORE LOSSES

Copper loss in each winding is calculated simply from the square of the rms random current times winding resistance as in the sinusoidal case.

Core-loss data for random fluxes do not exist. From the fact that the random flux reaches a peak value far less often than it does when the equivalent sine wave voltage and frequency (equations 27 or 42) are applied, one would expect core loss for the random case to be less than it is for the equivalent sinusoidal case. This was also verified experimentally by measuring the rate of core temperature rise of two identical transformers, one operated open circuit at its random noise rating and one at its equivalent sine wave rating. The sine wave transformer had a measurable and reasonable core temperature rise. The random-noise transformer, however, developed no perceptible rise in core temperature.

OTHER TRANSFORMER CONSTANTS

Depending upon how faithfully the power spectrum of the input voltage must be reproduced, and upon allowable phase shifts if the transformer is to be included in a feedback loop, the effects of magnetizing current, leakage inductances, and winding capacitances may be analyzed by the methods used for sine wave transformers operating over the frequency band of interest.

Possible Extensions

Preceding portions of this paper have outlined a basic method of computing the rms value of random flux in a transformer core, and have worked out two examples for rather simple random-noise voltage power spectra. This method can be extended to more complicated noise voltage power spectra. In doing this one may plot the integrand of equation 18 on a linear scale and integrate graphically to find the mean square value of random flux.

Alternatively, one might plot the voltage power spectrum in decibels on a logarithmic frequency scale, then approximate this curve by straight lines having slopes that are multiples of 12 decibels per octave. One could then set up integrals between the corner frequencies in a manner analogous to the steps taken from equations 30 to 42 and carry out the integrations analytically. The result would be an approximation that probably would be sufficient for practical purposes.

The general philosophy of this analysis might conceivably be profitably extended transformers for communication, speech, and music which are, in a sense, random signals. Some investigation would be necessary, however, to determine an appropriate power spectrum of the signal voltage and an appropriate probability density which will guide a decision as to maximum crest values of signal to be accommodated.

References

1. Random Processes in Automatic Control (book), J. Halcombe Laning, Jr., Richard H. Battin. McGraw-Hill Book Company, Inc., New York, N. Y., 1956.

2. Mathematical Analysis of Random Noise (book), S. O. Rice. In "Selected Papers on Noise and Stochastic Processes," Nelson Wax, editor. Dover Publications, Inc., New York, N. Y., 1954.

3. A Short Table of Integrals (book), B. O. Peirce. Ginn and Company, Boston, Mass., 1929.

Properties of Octafluorocyclobutane, a Dielectric Gas

F. W. BLODGETT
NONMEMBER AIEE

THERE IS a continuing need for improvement in all types of electric equipment. New applications call for higher operating temperatures, better reliability, smaller size, lighter weight, and little, if any, increase in cost. At the present level of design perfection, new materials having better properties are often required for making equipment which can handle higher voltages, or more power per pound or per cubic foot.

This paper describes the properties of a new dielectric gas. Enough data have already been obtained to show that octafluorocyclobutane, or C_4F_8, has a combination of properties making it very interesting for use in new electric equipment.

Electrical Properties

Octafluorocyclobutane, which is also called perfluorocyclobutane, is a cyclic fluorocarbon. Like other fluorocarbons, it is very stable, even at elevated temperatures. Also like other volatile fluorocarbons, C_4F_8 is an electronegative gas, i.e., it has a strong affinity for electrons. Its combination of electronegativity, high molecular weight, and chemical stability is responsible for its outstanding electrical properties.

In an electric field, C_4F_8 absorbs electrons to form ions of low mobility. This results in a fairly stable space charge which tends to make the electric field more uniform, and it greatly reduces the tendency of the electrons to form a conductive path. As a result, C_4F_8 has a high dielectric strength or breakdown voltage.

Breakdown voltage. Fig. 1(A) shows the breakdown voltage of C_4F_8 in a fairly uniform electric field. For easy reference, C_4F_8 is compared with air and with sulfur hexafluoride (SF_6), another electronegative gas which has already been used in some high-voltage electric equipment. As Fig. 1(A) shows, C_4F_8 has about the same dielectric strength as SF_6 at very close electrode spacing where the electric field is very nearly uniform. At wider spacings, where the field becomes less uniform, C_4F_8 will sustain 20% to 30% higher voltages than SF_6.

Nonuniform fields. The breakdown voltages of these gases are slightly less in a nonuniform than in a uniform field. Fig. 1(B) shows the higher dielectric strength of C_4F_8 as compared to SF_6. Although both electronegative gases are far superior to air, C_4F_8 sustains 30% to 50% higher voltages than SF_6.

Electrode geometry. The electrodes used to study dielectric strength in uniform electric fields were two chromium-plated disks, 3 inches in diameter, with suitably rounded edges. The set used for nonuniform fields was a 1/4-inch square aluminum rod and a 3-inch grounded disk. In both sets, electrode spacing was adjusted by means of a micrometer connected to one of the electrodes. The two sets of electrodes were used in a cylinder, with an 8-inch inside diameter, made of transparent plastic and filled with the gas to be tested.

Corona. There are no data for corona-starting voltages for C_4F_8, but an indication of the high voltages required for corona is given in Fig. 2, which compares the corona voltages for $n\text{-}C_4F_{10}$ with those for air in a nonuniform electric field. The gas $n\text{-}C_4F_{10}$ is the straight-chain fluorocarbon comparable to cyclic C_4F_8, and the electrical properties of these two fluorocarbon gases are very similar.

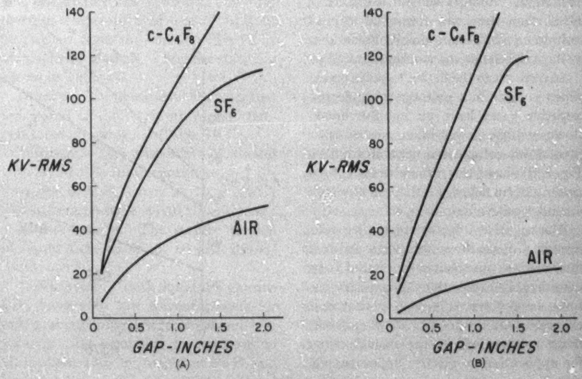


Fig. 1. Breakdown voltage

A—Uniform field
B—Nonuniform field

Paper 58-326, recommended by the AIEE Electrical Insulation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 13, 1958; made available for printing August 14, 1958.

F. W. Blodgett is with E. I. du Pont de Nemours and Company, Inc., Wilmington, Del.

The electrical data on C_4F_8 and the other gases shown in the figures of this paper were determined by Camilli, et al.,[1] and the authors take this opportunity to acknowledge their indebtedness to them.

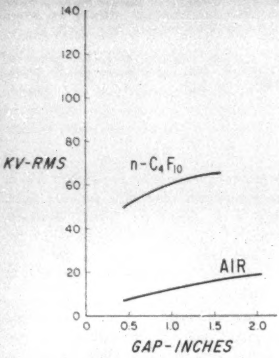


Fig. 2. **Corona voltage, nonuniform field**

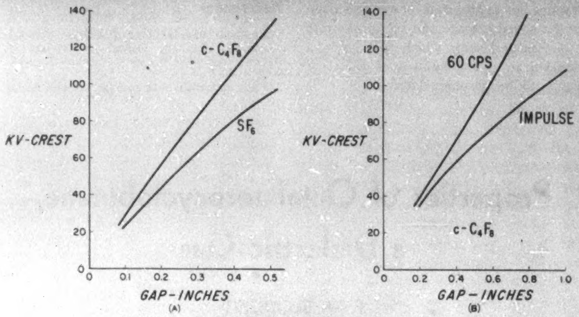


Fig. 3. **Impulse breakdown voltage**

A—Uniform field
B—Nonuniform field

C_4F_8 may be expected to have corona voltages 4 to 6 times higher than air, with the higher ratios in the more nonuniform fields. As yet the author has no data on the formation of any chemical breakdown products that may be formed from C_4F_8 during continuous corona operation.

Impulse. The behavior of dielectric gases under impulse voltages is important if the gas is to be used in equipment that can be subjected to impulses caused by switching in reactive circuits or by lightning.

Fig. 3(A) shows the breakdown voltages of C_4F_8 and SF_6, with the standard 1.5x40-microsecond positive impulse. In a uniform field, C_4F_8 has a higher impulse breakdown voltage than the 60-cycle breakdown voltages shown in Fig. 1(A), even when these are converted to crest voltages. C_4F_8 maintains its breakdown voltage difference, as compared to SF_6.

In contrast to the higher impulse breakdown voltage in a uniform field, electronegative gases have an impulse breakdown voltage lower than the 60-cycle breakdown voltage in a nonuniform field. Fig. 3(B) shows this difference for C_4F_8 in a nonuniform field, with the 1.5x40-microsecond positive impulse.

The negative impulse breakdown voltages in a nonuniform field for a series of fluorocarbon gases have been found to be considerably higher than the positive impulse breakdown voltages,[2] so that it is expected that the positive impulse breakdown voltages will be the critical values for application to electric apparatus design.

Pressure. The dielectric strength of all gases in a uniform electric field increases linearly with pressure, although the slopes and intercepts of different gases are different. These differences depend on the molecular structures of the gases in a way which is not yet understood.

Fig. 4(A) shows the breakdown voltages of C_4F_8, SF_6 and air at pressures up to 25 psig (pounds per square inch gage). The data for C_4F_8 were estimated from its breakdown voltage at atmospheric pressure and the slope of the breakdown voltage for its linear homolog, *n*-C_4F_{10}.

In a nonuniform field, on the other hand, corona and space charge effects distort the linear relationship. Fig. 5 shows that both the 60-cycle and the positive impulse breakdown voltages of C_4F_8 go through a maximum at relatively low pressures and that the impulse curve is almost flat. The author does not have data on SF_6 with the same set of electrodes. Therefore, if the field is nonuniform, there is no advantage in using an electronegative gas at a high pressure. As a matter of fact, in a closed system, careful control of the gas pressure by controlling temperature would be necessary to obtain reliable performance at the maximum dielectric strength rating.

Gas mixtures. Recent work at the General Electric Company[1,3] and the Westinghouse Electric Corporation[2,4] has shown how this anomalous behavior of electronegative gases can be controlled so that the high dielectric strengths of electronegative gases at high pressure can be used even in nonuniform electric fields. The ability of these electronegative gases to capture electrons and prevent break-

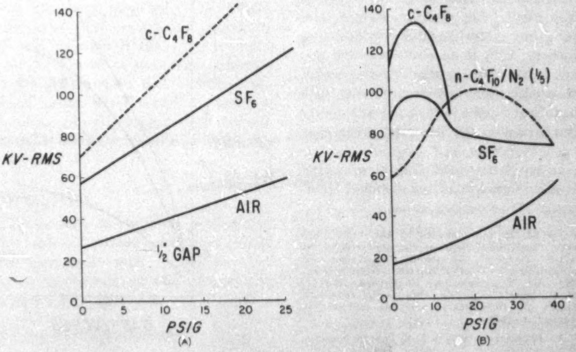


Fig. 4. **Breakdown voltage under pressure**

A—Uniform field
B—Nonuniform field, 1-inch gap

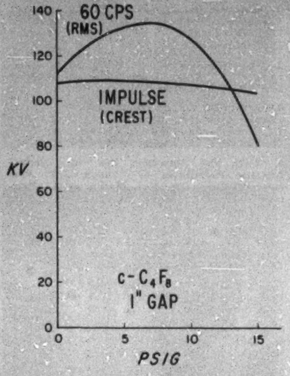


Fig. 5. Positive impulse breakdown voltage under pressure, nonuniform field

down is so great that they have a strong effect even when diluted with nitrogen. The nitrogen increases the pressure range over which the mixture shows increasing dielectric strength with increasing pressure. This effect is shown in Fig. 4(B) with n-C_4F_{10}.

In a nonuniform field, the dielectric strengths of C_4F_8 and SF_6 reach maximum values at rather low pressures. But a mixture of n-C_4F_{10} and nitrogen, at a ratio of one part n-C_4F_{10} to five parts of nitrogen by volume, reaches a high dielectric strength at moderate pressures and has a somewhat broader maximum, so that accurate pressure control is not so critical as with the pure gas. Although a patent claiming this same effect with C_4F_8 has been issued,[3] no data obtained with this particular set of electrodes have been published.

Table I. Physical Properties of c-C_4F_8

Molecular weight	200.04	200.04
Boiling point	21.1 F	−6.04 C
Freezing point	−42.5 F	−41.4 C
Critical temperature	239.5 F	115.3 C
Critical pressure	393 psia*	26.7 atm†
Critical density	38.2 lb/cu ft	0.612g/ml‡

* Psi absolute.

† Atmosphere.

‡ Milliliter.

Table II. Vapor Pressure of C_4F_8

Temperature		Pressure, Psig
F	C	
0	−18	−5.9
32	0	4.1
70	21	25.4
100	38	52.1
150	66	125.3

Depending on the operating temperature range and the uniformity of the electric field, these mixtures of electronegative gases with nitrogen may show a cost advantage over the pure gas at comparable dielectric strengths. The mixtures also permit operation of a dry gas system at higher pressures and lower temperatures than does the pure electronegative gas, because the partial pressure of the condensable gas remains relatively low.

Physical Properties

C_4F_8 is a dense, colorless, odorless gas at atmospheric temperature and pressure. Its molecular weight, boiling and freezing points, and critical data are shown in Table I.[5]

Since C_4F_8 has a fairly high boiling point its vapor pressure is moderate at room temperature. Table II shows the vapor pressure,[5,6] over a range of temperatures. Although this limits the temperature range at which pure C_4F_8 may be used in dry dielectric gas-filled apparatus, it makes it practical for evaporative cooling applications. Its chemical inertness and its very low solvent power for magnet wire insulation and other solid insulation means that liquid C_4F_8 can be sprayed onto transformer coils in a closed system where it will remove heat by evaporation, and then condense on the shell, giving up the heat to the atmosphere. C_4F_8 has a heat of vaporization of 49.8 Btu/lb (pound) or 28 cal/g (calories/gram) at its normal boiling point.

Gas mixtures. When one volume of C_4F_8 is diluted with four volumes of nitrogen, the mixture can be used as a dry gas down to temperatures of −12 C (10 F) at 40 psig, or down to −38 C (−36 F) at atmospheric pressure. Further data are shown in Table III.

Density. The densities of the saturated vapor and of the liquid at various temperatures are shown in Table IV.[5]

Viscosity. The viscosity of liquid C_4F_8 at 70 F is 0.455 centipoise and the viscosity of C_4F_8 vapor at 70 F and 1 atmosphere pressure is 0.0117 centipoise.[7]

Surface tension. The surface tension of liquid C_4F_8 at 68 F is 8.41 dynes/centimeter.[8]

Solvent power. Both liquid and gaseous C_4F_8 have very low solvent powers for oils, rubber compounds, and magnet wire enamels. Its solubility in refrigeration oil is about 15% by weight at 68 F, and the solubility of oil in liquid C_4F_8 is about 2% by weight at 68 F. So little C_4F_8 dissolves in rubber that it does not leak from a rubber balloon. After immersion in liquid C_4F_8, the tensile strength of a neoprene compound was 85% of its original value.[7]

Table III. Saturation Temperatures of C_4F_8/Nitrogen Mixtures

Pressure, Psig	Saturation Temperature, C			
	0*	1*	2*	4*
0	−6	−21	−29	−38
5	1	−15	−23	−33
10	7	−10	−18	−29
15	13	−5	−15	−25
20	17	−1	−11	−22
25	21	2	−8	−19
30	25	5	−5	−16
35	28	8	−2	−14
40	31	11	0	−12

* Volume of nitrogen/volume of C_4F_8.

Table IV. Density of Saturated C_4F_8

Temperature, C	Vapor, Lbs/Cu Ft	Liquid, Lbs/Cu Ft	Pressure, Psia
−17.8	0.365	104.0	8.8
0	0.710	99.7	18.8
21.1	1.505	94.9	40.1
37.8	2.52	90.3	66.8
65.6	5.45	81.3	140.0

Structure and polarization. The detailed structure of C_4F_8 has been established by electron diffraction[9] and infrared absorption.[10] As would be expected from its symmetry as a ring compound, C_4F_8 has a zero dipole moment.[11] This means that its electrical loss factor is very low. Because of the high electronegativity of fluorine, the molar refraction of C_4F_8 is 19.2, the molar polarization is 24.8, the electronic polarization is about 18.5, and the atomic polarization is 6.3.[11]

Thermodynamic properties. Data for the thermodynamic properties of C_4F_8 have been obtained,[5] and values of the important functions are available.

Chemical Properties

C_4F_8 is a saturated cyclic fluorocarbon having four CF_2 groups connected in a ring. Like other saturated fluorocarbons, it is very stable chemically. Like other electronegative gases containing halogen atoms, C_4F_8 is ionized by high-energy electrons, losing a fluorocarbon ion, rather than an electron, to form a heavy positive ion. The low mobilities of these ions in an electric field are important factors contributing to its high dielectric strength. Its ionization potential is about 12.8 electron volts.[7]

Chemical reactions. C_4F_8 is highly stable, having a heat of formation of 352 kcal/mole.[12] It does not react with other materials under ordinary chemical conditions. At high temperatures, it dissociates to form carbon tetrafluoride and carbon.[7] The rate of this reaction has not been measured accurately, but it becomes rapid only above 600 C unless catalysts are present. Other products of high-temperature dissociation of C_4F_8 include toxic unsaturated fluorocarbons, such as tetrafluoroethylene[7] and octafluoroisobutylene.[13] Nickel and Inconel are inert, so C_4F_8 may be used continuously with these metals at temperatures up to 400 C. Steel, bronze, and yellow brass catalyze the decomposition and should not be used at temperatures above about 250 C. Steel and copper test pieces sealed in glass tubes with refrigeration oil and C_4F_8 caused no decomposition and showed no copper plating or corrosion during 500 days at 150 C. Copper with C_4F_8 alone does not tarnish at 232 C. C_4F_8 stability tests are still in progress.

Flammability. C_4F_8 was tested in a combustion buret. It is nonflammable in all proportions in air, but some mixtures with pure oxygen can be ignited by a spark. The products are carbon tetrafluoride, carbon dioxide and carbonyl fluoride.[7] This behavior is similar to that of other fluorogases.

When 10 pounds of C_4F_8 was mixed with the air in a 1,000-cu-ft (cubic-foot) test room, and a 2-burner gas stove was started, analysis of the well-mixed air showed 1.08% carbon dioxide, 0.038% carbon monoxide, 0.79% hydrogen fluoride, and 0.016% carbonyl fluoride after 30 minutes' burning time. This indicates the hazard of breathing mixtures of C_4F_8 and air that have been exposed to arcs or flames. Since hydrogen fluoride is very irritating in small concentrations, there is ample warning of the hazard.

Hydrolysis. C_4F_8 is extremely stable to hydrolysis. Neutral, alkaline, and acidic aqueous solutions saturated with C_4F_8 at temperatures up to 130 F (54 C) in lacquer-lined tin-plated steel cans for one year showed less than one part per million fluoride ion concentration in the aqueous phase.[7]

Toxicity

Preliminary work shows that the toxicity of C_4F_8 is very low, but adequate ventilation should be used in any experimental work until the results of further toxicity tests, which are now in progress, are available. Breathing C_4F_8 that has been exposed to flame temperatures should always be avoided, since under some conditions it may decompose slightly and contain toxic or irritating decomposition products. Samples of C_4F_8 or Freon-C318,[*] are now available for evaluation as a dielectric gas.

References

1. The Dielectric Behavior of Some Fluorogases and Their Mixtures, **G. Camilli, T. W. Liao, R. E. Plump.** *AIEE Transactions*, vol. 74, pt. I, Nov. 1955, pp. 637–42.

2. Electric Breakdown of Perfluorocarbon Vapors and Their Mixtures with Nitrogen, **L. J. Berberich, C. N. Works, E. W. Lindsay.** *Ibid.*, pp. 660–66.

3. Improvements in and Relating to Electrical Apparatus, **General Electric Company, Ltd.** *British Patent no. 782,130*, 1957.

4. Improvements in or Relating to Electrical Apparatus Provided with Gaseous Insulating Atmospheres, **L. J. Berberich, C. N. Works.** *British Patent no. 671,780*, 1952.

5. Data and Equations for the Thermodynamic Properties of "Freon-C318" Perfluorocyclobutane, **J. J. Martin, et al.** Engineering Research Institute, University of Michigan, Ann Arbor, Mich. (unpublished).

6. Vapor-Liquid Equilibria of Some Fluorinated Hydrocarbon Systems, **G. H. Whipple.** *Industrial and Engineering Chemistry*, Washington, D. C., vol. 44, July 1952, pp. 1664–67.

7. Properties of "Freon-C318." E. I. du Pont de Nemours & Company, Inc., Wilmington, Del. (unpublished).

8. Properties of Octafluorocyclobutane, **G. Bambach.** *Kältetechnik*, Karlsruhe, Western Germany, vol. 8, 1956, pp. 334–39.

9. The Molecular Structures of Octafluorocyclobutane and of Methylcyclobutane, **H. P. Lemaire, R. L. Livingston.** *Journal of the American Chemical Society*, Washington, D. C., vol. 74, Nov. 1952, pp. 5732–36.

10. Spectroscopic Properties of Fluorocarbons and Fluorinated Hydrocarbons. *NRL Report 3567*, Naval Research Laboratory, Washington, D. C., 1949.

11. The Dipole Moments and Molecular Structures of Some Highly Fluorinated Hydrocarbons and Ethers, **A. di Giacomo, C. P. Smyth.** *Journal of the American Chemical Society*, Washington, D. C., vol. 77, Feb. 1955, pp. 774–77.

12. Thermochemical Studies on Fluorocarbons, **H. C. Duus.** *Industrial and Engineering Chemistry*, Washington, D. C., vol. 47, July 1955, pp. 1445–49.

13. Perfluoroisobutene, **W. H. Pearlson, L. J. Hals.** *U. S. Patent no. 2,617,836*, 1952.

14. Octafluorocyclobutane, **F. B. Downing, A. F. Benning, R. C. McHarness.** *U. S. Patent no. 2,384,821*, 1945.

15. Octafluorocyclobutane, **J. Harmon.** *U. S. Patent no. 2,404,374*, 1946. Reissue *23,425*, 1951.

---

[*] Trade-mark for octafluorocyclobutane.[14, 15] Requests for samples or for further technical data should be addressed to "Freon" Products Division, E. I. du Pont de Nemours and Company, Inc., Wilmington 98, Del.

Magnet Wire for 600 C Temperature Transformers

W. W. WAREHAM
NONMEMBER AIEE

AS A PART OF WORK performed on a government contract, magnet-wire materials were investigated for application in electronic power transformers and inductors. One of the most critical environments specified for the application was 1,000-hour life in the ambient temperature range −65 C (degrees centigrade) to 500 C. This paper summarizes the planning and results of an ultrahigh-temperature heat-aging program designed to screen selectively the most promising magnet wires to meet this requirement.

Planning the Heat-Aging Program

An ambient temperature requirement of 500 C meant that the magnet wire would be exposed to even higher temperatures due to internal heat rise and hot-spot conditions. The actual requirement in terms of magnet-wire performance was therefore arbitrarily selected as 1,000 hours at 600 C. The selection of materials to be evaluated, the properties to be measured, and the type of test fixture to use were challenging questions which had to be answered before the program could be started.

SELECTION OF MAGNET WIRES

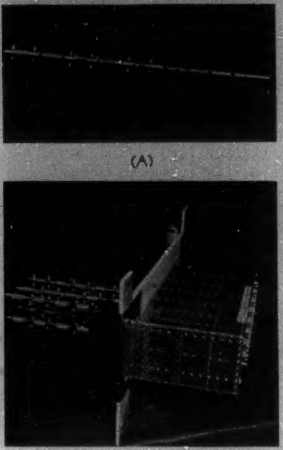
A survey of available magnet wires which had potentiality for performing 1,000 hours at 600 C revealed that the choices were quite limited. It was obvious that copper could not be used directly at this temperature, so the selection of the magnet wires involved both

---

Paper 58-117, recommended by the AIEE Electrical Insulation Committee (formerly the Dielectrics Committee) and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 14, 1957; made available for printing November 17, 1958.

W. W. Wareham is with the General Electric Company, Fort Wayne, Ind.

This work was supported by the U. S. Air Force through its project engineer, J. F. Rippin, Jr., under Contract no. *AF33(616)3623*. The author also wishes to acknowledge the assistance of A. M. Greider, G. H. Bartel, and G. Walters.


(A)

(B)

Fig. 1. Samples for wire-evaluation test

A—Specimen
B—Rack

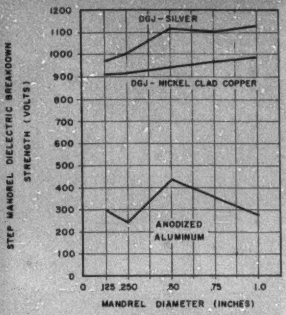


Fig. 2. Initial step mandrel dielectric strength as a function of mandrel diameter

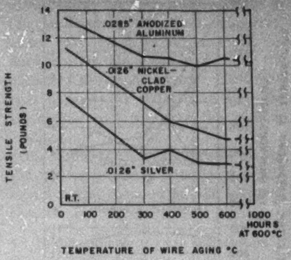


Fig. 3. Test wire tensile strengths after a 40-hour exposure to several elevated temperatures. All measurements taken at room temperature immediately after cooling

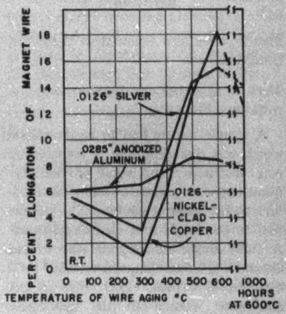


Fig. 4. Per-cent magnet wire elongation after a 40-hour exposure to several elevated temperatures. All measurements taken at room temperature immediately after cooling

the insulation and conductor materials. Much interest was focused on the possibility of using a ceramic-coated conductor but most of these materials were still in the development stage and available only in limited quantities. The major concern was to find a magnet wire that could be supplied in quantity by well-known manufacturing processes. Many insulations were considered but were rejected because of insufficient physical properties for winding, and some possessed inferior electrical properties at 600 C. The three samples which were selected for complete evaluation were:

1. A 0.0126-inch double glass-served silicone-varnish-treated, 27% nickel-clad copper wire.

2. A 0.0126-inch double glass-served silicone-varnish-treated silver wire.

3. A developmental sample of 0.0285-inch anodized aluminum wire. Although this wire is not available commercially, its excellent high-temperature properties, low film build, and long-range economic aspects made it a desirable candidate.

PROPERTIES TO BE MEASURED

Magnet-wire properties are usually investigated in terms of physical strength first and electrical and chemical properties later. This is true because magnet wire must be wound into a coil shape before it ever begins to perform its electrical function. Physical properties and nondestructive electrical properties were emphasized in the initial tests at room temperature.

Initial Tests at Room Temperature

Magnet-wire samples were tested for the following characteristics:

1. Flexibility per-cent elongation to crack insulation.

2. Step mandrel dielectric strength.

3. General Electric Repeated Scrape Abrasion Test.

4. Film build and concentricity of insulation.

5. Insulation resistance using a 35-volt d-c megohmmeter.

6. Twist dielectric strength.

7. Tensile strength, conductivity, and per-cent elongation of conductors.

Tests at Elevated Temperatures

A conditioning period of 40 hours at temperatures of 300 C, 400 C, 500 C, and 600 C was established to remove the volatile components of the wire insulation. New samples were used at each test temperature. Values of insulation resistance and dielectric strength were determined after each of the conditioning periods. These same readings were taken on another set of wires aged for 1,000 hours at 600 C. The tensile, elongation, and conductivity characteristics of the con-

Table I. Physical Test Results

Initial Test Data

Wire Title	Mandrel Flexibility	Per-Cent Elongation to Crack Insulation	Twist Dielectric Strength, Volts	Scrape Abrasion Resistance Strokes, 200-Gram Load	Film Build, Mils
0.0126-inch DGJ nickel-clad copper*	1X, outer layer opens; inner layer, OK; 3X, OK	4.2: wire breaks, no visual damage to insulation	1,500	6	7.5
0.0126-inch DGJ silver*	1X, outer layer opens; inner layer, OK; 3X, OK	5.6: wire breaks, no visual damage to insulation	1,600	6	7.4
0.0285-inch anodized aluminum	0.300 inch, slight cracking; 0.400 inch, OK, no visual cracking	3: cracking 1: slight cracking 0.5: OK 6: wire breaks	478	13	0.7

* Double, glass-served, silicone-varnish-treated.

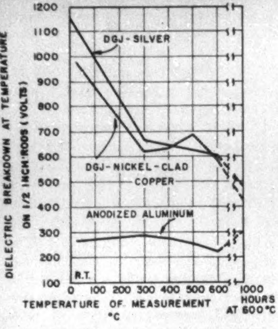


Fig. 5. Dielectric strength measured hot after a 40-hour exposure to several temperature levels

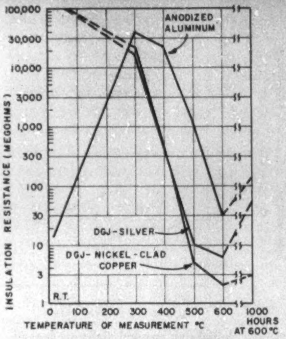


Fig. 6. Wire insulation resistance after a 40-hour exposure at several temperature levels

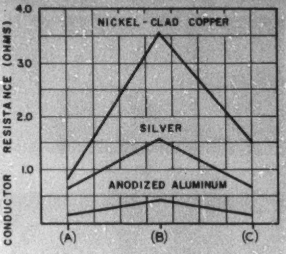


Fig. 7. Changes in conductor resistance due to aging at 600 C

A—Initial resistance of coil at room temperature

B—Coil resistance at 600 C after 1,000 hours at 600 C

C—Coil resistance at room temperature after 1,000 hours at 600 C

ductors were also followed at these test conditions.

Type of Electrical Measurement Sample Fixture

A sample fixture had to be developed which was representative of transformer coil construction and would provide adequate support for the wire and allow readings to be taken at 600 C. The sample had to be simple and easily replicated. A tight single strand coil was wound around a 1/2-inch-diameter mandrel for a distance of 1 inch. Both ends of the coil were secured to the cylindrical mandrel by threading through ceramic bushings mounted in sets of two which were staggered 90 degrees alternately along the mandrel. Five such coils were mounted on a single 18-inch-long 304 stainless steel rod 1/2 inch in diameter which had been given a coat of flash nickel and then rhodium plated to protect against surface oxidation. The five coils were wound close together covering a distance of 10 inches along the rod. The next 2 inches of the rod were left open to provide space between a high-temperature oven door and the nearest coil. The remaining 6 inches of the rod protruded out through the door to be used for mechanical support and electrical connections. The leads from each of the five coils inside the oven were guided through the door and test rack by long ceramic tubes. Five coils wound on a single rod can be seen in Fig. 1(B).

Electrical measurements were taken between each of the five coils and the stainless steel mandrel. Two rods containing samples of each test wire were made for every temperature condition so that an average of ten readings could be obtained.

Discussion of Results

A comparison of the physical properties of the test specimens can be found in Table I. All the samples are deficient in insulation flexibility and abrasion resistance. However, all the wires could probably be used in transformer construction with proper care in handling.

Fig. 2 summarizes the initial step mandrel dielectric strength values as a function of mandrel diameter. The glass-served wire samples give by far the most protection even in the 0.125-inch mandrel. The anodized aluminum was lower but in terms of film build (volts per mil) it ranked the highest.

Figs. 3 and 4 show how heat exposure will change the tensile strength and percent elongation qualities of the three conductors. The wires illustrated an annealed effect with lower tensile and increased elongation.

Fig. 5 records the dielectric strength readings of the three wire samples which had been wound around the 1/2-inch-diameter rods and exposed to 300 C, 400 C, 500 C, and 600 C. The glass-served wires dropped from approximately 1,000 volts at room temperature to 600 volts at the four high-temperature levels. After 1,000 hours at 600 C the glass-served wires were still showing about 450 volts. The anodized aluminum wire remained quite constant at about 250 volts for all temperatues and improved to 300 volts after 1,000 hours at 600 C.

Some unusual insulation-resistance properties were observed from the data summarized in Fig. 6. The glass-served wires dropped from 250,000 megohms insulation resistance at room temperature to just a few megohms at 600 C. The glass-over-silver wire turned pale yellow after 1,000 hours at 600 C and an increase in insulation resistance was noted. The porous nature of the anodized aluminum wire and its ability to hold moisture

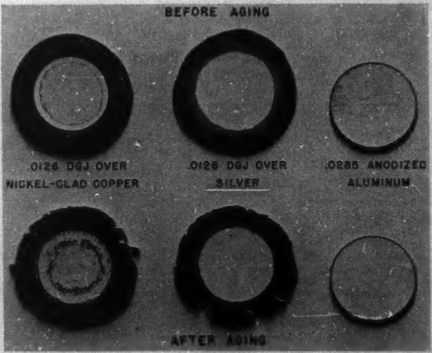
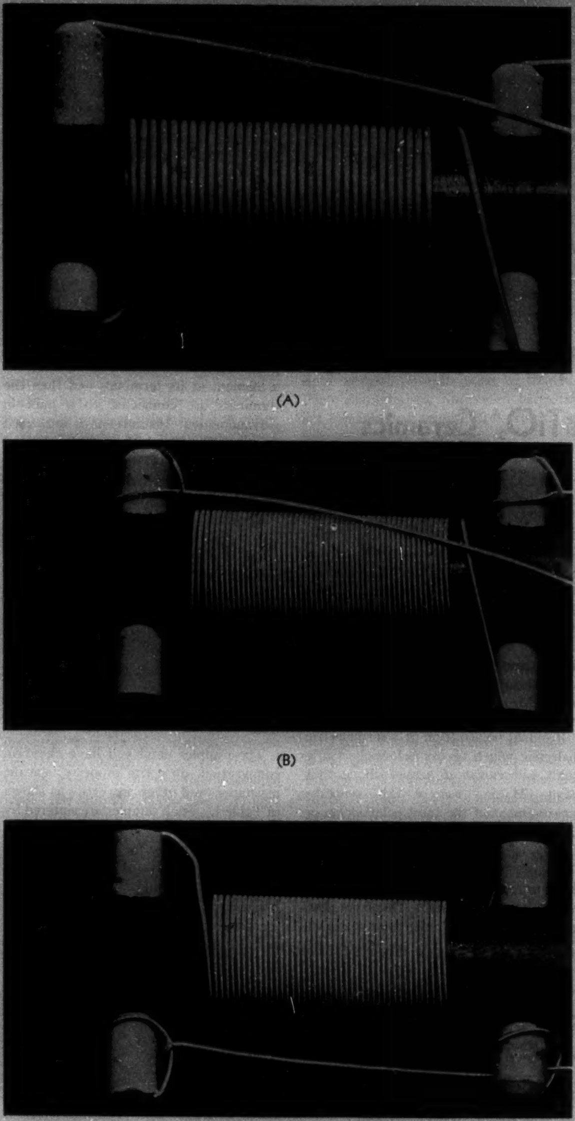


Fig. 8. Magnet wire cross sections before and after 1,000 hours at 600 C



(A)

(B)

(C)

Fig. 9. Wires aged 1,000 hours at 600 C

A—0.0285-inch anodized aluminum
B—0.0126-inch double, glass-served silver
C—0.0126-inch double, glass-served nickel-clad copper

caused rather low and erratic readings in the "as received" condition. However, added heating and drying improved the insulation resistance and it remained the highest of all samples tested. The 1,000-hour exposure at 600 C improved the insulation resistance over the 40-hour readings.

Fig. 7 shows the resistance values for the various conductors wound on a 1/2-inch-diameter mandrel for the following conditions:

1. Initial readings at room temperature.
2. Readings after 1,000 hours at 600 C taken at 600 C.
3. Readings after 1,000 hours at 600 C taken at room temperature.

All the conductors showed an increase in resistance at the end of 1,000 hours at 600 C except anodized aluminum. The nickel-clad conductor showed a serious increase in resistance. An explanation for this increase can be seen in Fig. 8, where the nickel has migrated into the copper forming a high-resistance alloy layer.

DESCRIPTION OF SAMPLES AFTER AGING

Fig. 9(A) shows the physical appearance of the anodized aluminum after 1,000 hours at 600 C. Small corrosion spots can be seen along the surface of the wire. The exact cause of these spots is unknown but it seems reasonable to assume that they may have been caused by remnants of corrosive electrolyte from the anodizing process. If this is the cause then it is important for all anodized wire to be thoroughly washed and cleaned before it is subjected to long exposures at high temperature.

Figs. 9(B) and 9(C) show how well the glass-served wire remained in good condition after 1,000 hours at 600 C. The glass fibers were extremely dry and brittle and could be scraped off easily but they remained in place to provide a consistent space factor.

All the photographs show a decided darkening of the rhodium plating which was used to coat the *304* stainless steel rods. In spite of the oxidation resistance and inertness of rhodium the severity of the test temperature and time did create an insulating film over the electrode. In order to determine the magnitude of this oxide coating, the layers of wire were carefully removed from the rods and replaced turn-by-turn with 0.010 bare copper wire. The insulation resistance between this wire and the rod averaged 0.6 ohm which indicated a very small barrier. As a second test a 1/4-inch strip of aluminum foil was wrapped around the rod at the end where no wire coil had been wound. The insulation resistance measured approximately 100,000 ohms at room temperature showing that the oxide film was much heavier where there had been no wire to protect the rod. This experience suggests that a platinum-clad electrode would be necessary to eliminate completely the interference of electrode oxide for conditions near 1,000 hours at 600 C or higher. Fig. 8 shows a composite of all the magnet-wire cross sections before and after the 1,000-hour exposure to 600 C. The nickel-clad copper formed a high-resistance alloy layer at the nickel-copper interface. Once the silicone varnish has been burned out of the glass-serving, the glass becomes loose and fragile as shown in the lower cross sections. Little change can be seen in the silver and aluminum conductors.

Conclusions

The results of this study have led to the following conclusions:

1. The 0.0126-inch double glass-served silicone-varnish-treated silver wire performed the best in most of the evaluation tests.

2. The 0.0126-inch double glass-served silicone-varnish-treated nickel-clad copper performed well except for the increase in conductor resistance brought about by a nickel-copper interaction after 1,000 hours at 600 C.

3. Anodized aluminum showed promise in most of the tests, especially when hot. This would be the recommended wire if low film builds become a necessity.

On the Anomaly in Residual Polarization of $BaTiO_3$ Ceramics

Y. SAITO
MEMBER AIEE

S. YAMANAKA
NONMEMBER AIEE

SPONTANEOUS polarization of barium-titanate ($BaTiO_3$) ceramics has been noted in research for its ferroelectricity and for its practical use. It is natural that a great deal of interest has been attached to spontaneous polarization as the property is quite essential in the ferroelectric phenomena. Nevertheless when measurements of residual polarization of $BaTiO_3$ were being made, the authors found that a very large amount of polarization occurred at times and it could not be attributed to spontaneous polarization in any case.[1] Blood and his co-workers[2] also found large residual polarization more than 150 μcoul/cm^2 (microcoulomb/centimeter^2) in $BaTiO_3$ ceramics accompanied by the anomalous phenomena in piezoelectric oscillation. After examining the phenomena it was recognized that some sort of polarization should be superposed upon the proper spontaneous polarization. As such phenomena are never favorable to the practical use of ferroelectric substance, investigations have been made especially to find a way to diminish their effect. Thus it was found that the mixing ratio of raw materials of the ceramics seemed to have a remarkable relation to the anomalous polarization, and the reduction of $BaTiO_3$ seemed to exist in accompaniment with such phenomena. The result obtained is reported in this paper.

Measurement of Residual Polarization and Samples Prepared for Measurement

Residual polarization of a ferroelectric dielectric which occurs below the Curie point will vanish if heated above that point. Making use of this property measurement of the total polarization of the dielectric is possible. Small plates of $BaTiO_3$ ceramics are used as specimens; they are glazed with silver and then soldered to terminal wires. Thus small plate condensers are formed. These are first heated above the Curie point to erase their history, then the condensers are polarized under various conditions. A polarized condenser is short-circuited for a moment at its terminals and then the terminals are connected to a ballistic galvanometer. Instantly it is immersed in a paraffin oil which has been heated beforehand at 170 C (degrees centigrade). Then the electric charge passing through the galvanometer, which is identical with the residual polarization of the specimen, is observed. However, the amount thus obtained contains some sorts of polarization and has comparatively large relaxation time in addition to the proper spontaneous polarization. For the purpose of abbreviation the symbols P_r, P_s, and P_t are used for total polarization, spontaneous polarization, and the difference between them respectively. We may sometimes call P_t anomalous polarization.

The samples employed have chemical composition, dielectric property, specific resistance, and other properties as shown in Table I. Such samples are sintered by duplicate heating. In the first stage, ingredients are mixed for 10 hours in a pot mill and then cakes are made from the mixture by pressing. These cakes are baked at 1,200 C for an hour. In the second stage the cakes are crushed in a pot mill again, then formed by pressing,

Table I

No.	Mixing Ratio $BaCO_3$:TiO_2	Curie Point, C	20 C	Curie Point	tan (10^{-4})	$(10^{10}$ cm)	(μc/cm^2)	Note
101	1:1	112	1,350	6,300	250	30	10.4	
102	1:1	108	1,300	2,000	800	40	0	
103	1.01:1	109	1,500	6,600	380	50	3.6	
104	1:1.01	109	1,150	6,800	440	90	13	
105	1.03:1	103	2,300	5,000	400	155	0	
106	1:1.03	113	750	4,000	450	5	55.6	
107	1:1	112	2,050	8,000	490	20	4	No. 102 ground and baked again
108	1:1	111	1,400	6,000	440	65	13.2	Previous baking 4 hours
109	1:1	111	1,370	6,400	530	70	8.8	Mixing 50 hours
110	1:1					45	6.8	1,450 C 2 hours baking

Paper 58-879, recommended by the AIEE Electrical Insulation Committee (formerly the Dielectrics Committee) and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22-27, 1958. Manuscript submitted December 20, 1957; made available for printing October 22, 1958.

Y. SAITO and S. YAMANAKA are with the Tokyo Institute of Technology, Tokyo, Japan.

Vol.
Pt. 1
195

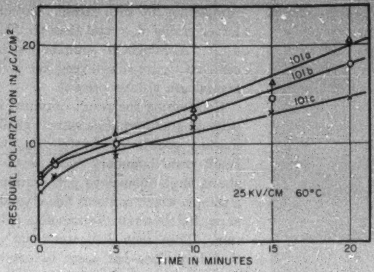


Fig. 1. Relation between residual polarization and time-of-voltage impression in several specimens

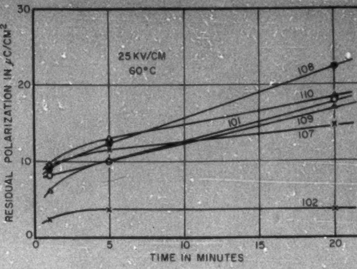


Fig. 2. Relation between residual polarization and time-of-voltage impression in $BaTiO_3$ ceramics prepared by various processes

and finally sintered at 1,400 C for an hour.

Dielectric properties are measured by the bridge method using high-frequency electric source of 1 kc. Specific resistance is measured by galvanometer deflection applying a direct potential of 100 volts on each specimen. As to the anomalous polarization P_i, shown in Table I, an explanation will be given later on.

Each specimen is a disk 10 mm (millimeter) in diameter and 0.5–1 mm in thickness. They are first silvered and then lead wires are soldered on both surfaces of the disks.

Factors Affecting Anomalous Polarization

Before examining a large number of measured samples, the relation between the residual charge P_r and the time-of-voltage impression t is precisely examined with a few samples. Comparatively high voltage is applied on each sample for 10 seconds, 1, 5, 10, 15, and 20 minutes successively. Examples of the relation between P_r and t are shown in Fig. 1. It is clear that the relations show straight lines on the major part of the curve except near its origin. Accordingly a few examples will be given for the purpose of determining the position of such a straight line since so many samples must be treated. Each characteristic illustrated hereafter is a typical result selected from the measurements on 5–10 samples.

The relation between P_r and t is measured with samples shown in Table I. Each measurement is always carried out at 60 C impressing 25 kv/cm. Thus the effects of the difference of manufacturing process, mixing proportion, porosity, and grain structure on the anomalous polarization of samples are examined. For example, as the spontaneous polarization of $BaTiO_3$ is believed to be a few μc/cm², such a sample as no. *108* in Fig. 2, which has a large value of P_r, must be considered to have large anomalous polarization. But no. *102* does not seem to have so large anomalous polarization. The values of P_i in Table I are calculated as follows: By extending the straight-line portion of the curve which shows the relation between P_r and t in Fig. 3, P_{r0} is obtained from the intersection of the straight line with the ordinate. P_i is calculated by substructing P_{r0} from the value of P_r at 20 minutes on the abcissa.

Effect of Manufacturing Process

Fig. 2 shows the results of various samples of different manufacturing processes, in which no. *109* has a very long mixing time of 50 hours, no. *108* has a very long baking time of 4 hours, and no. *110* is sintered for 2 hours at 1,450 C. Such variety in manufacturing process does not seem to cause any substantial difference on anomalous polarization. No. *102* does not seem to show any anomalous phenomenon. This sample is immediately sintered at 1,400 C without going through the previous baking.

Porosity

Since ceramics are generally very porous the effect of the internal surface of the pores on anomalous polarization is measured. The sample no. *102* is much more porous than no. *101* as shown in Fig. 4. Nevertheless, no. *102* does not seem to have any anomalous polarization as indicated in Fig. 2. Therefore porosity may have no direct effect on anomalous polarization.

Mixing Proportion

The mixing proportion of $BaCO_3$ and TiO_2 seems to have a considerable effect on the anomalous polarization. No. *104* and no. *106* in Fig. 5 which have excess TiO_2 amounting to 1 and 3 mole per cent respectively show large anomalous polar-

Table II.

No.	Composition	Curie Temperature, C	20 C	Curie Temperature	tan (10^{-4})	(10^{10} cm)	(μc/cm²)
201	(95Ba-5Sr)TiO_3	98	2,000	8,400	300	115	10.2
202	(95Ba-5Ca)TiO_3	113	1,220	7,600	190	170	1.3
203	(95Ba-5Mg)TiO_3	94	1,340	5,800	450	35	40.8
204	(95Ba-5Pb)TiO_3	132	900	6,900	250	195	2.8
205	Ba(95Ti-5Zr)O_3	96	1,700	8,650	470	98	5
206	Ba(95Ti-5Sn)O_3	83	3,400	9,900	580	12	5.8
207	BaTiO+0.3% Fe_2O_3	82	1,700	8,300	350	11	10.6
208	$BaTiO_3$+3%Fe_2O_3	71	1,800	2,250	2,500	1.8	0.4
209	$BaTiO_3$+2%$(NH_4)_2HPO_4$	117	880	3,600	500	30	9.6
210	203+3% MgO					220	1.9

The proportion of Fe_2O_3 and (NH_4) HPO_4 is weight per cent.

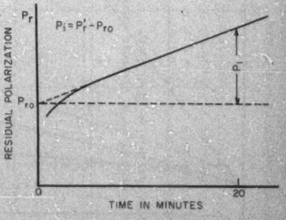


Fig. 3. Calculating method of P_i in Tables I and II

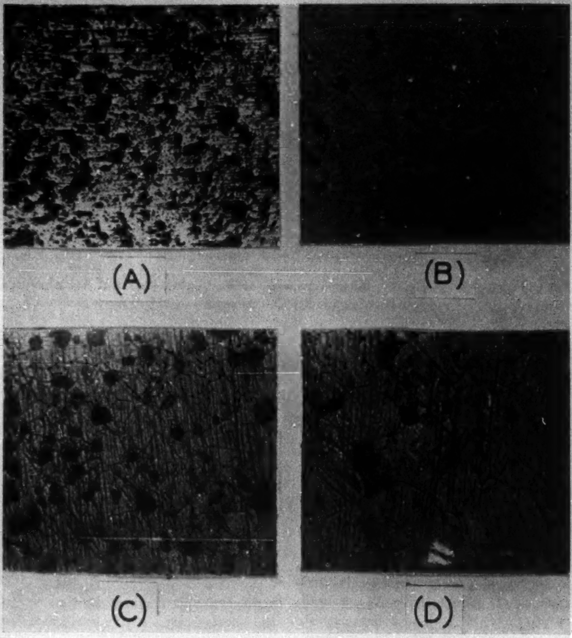


Fig. 4. Photomicrographs of unetched samples and etched samples

A—No. 102, ×33
B—No. 101, ×93
C—No. 101, ×133
D—No. 106, ×133

ization. However no. *103* and no. *108* which have excess $BaCO_3$ amounting to 1 and 3 mole per cent respectively show very small anomalous polarization.

GRAIN STRUCTURE

Applying metallurgical techniques to the polish and corrosion of ceramics, Kulcsar[3] observed the crystal structure of the ceramics. He reported that considerable growth of crystal had been observed in a sample having excess TiO_2 and just the reverse in a sample containing excess $BaCO_3$. In this paper a number of samples are examined by the same method as Kulcsar's. As shown in Fig. 4(D), sample no. *106* with excess TiO_2 contains larger grains of crystal than no. *101*, Fig. 4(C). Although no samples having excess $BaCO_3$ are shown very fine grains of crystal are always present.

Considering the result obtained in the preceding paragraph, a sample exhibiting marked growth of crystal, i.e., having small grain boundary such as no. *106*, shows large anomalous polarization, and a sample which contains fine crystal such as no. *105* shows small anomalous polarization. This fact indicates that grain boundary does not seem to contribute much to the anomaly; i.e., free ions wandering through the grain boundary do not seem to play any leading role in anomalous polarization.

Nature of Anomalous Polarization

As it is not easy to separate P_i (anomalous polarization) from P_s (spontaneous polarization) definitely it is difficult to know the nature of P_i accurately. However, as the spontaneous polarization of $BaTiO_3$ ceramics is believed to be a few $\mu c/cm^2$ the amount of P_i can be derived approximately as follows.

TEMPERATURE CHARACTERISTIC OF ANOMALOUS POLARIZATION

Fig. 6 shows the change of P_r of $BaTiO_3$ referring to the time-of-voltage impression with ambient temperature as a parameter. From these results it is easily understood that the higher the temperature rises the larger the anomalous polarization grows. In Fig. 7 the ordinate represents the logarithm of the inclination of straight-line portion of each curve in Fig. 6, and the abscissa shows the parameter of the curves in Fig. 6 measured by absolute scale. The relation is shown fairly well as a straight line.

Next, the characteristics of P_r near the Curie point are examined. As the Curie

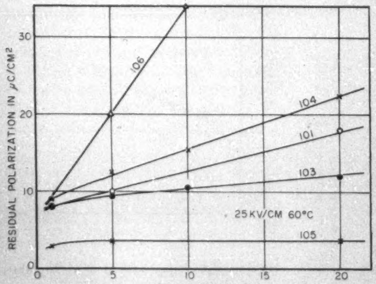
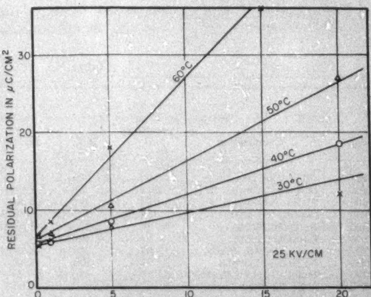


Fig. 5 (left). Relation between residual polarization and time-of-voltage impression in $BaTiO_3$ ceramics with various mixing proportions of $BaCO_3$ and TiO_2

Fig. 6. Relation between residual polarization and time-of-voltage impression in $BaTiO_3$ ceramics at various temperatures



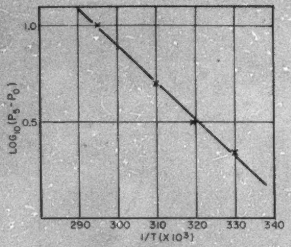


Fig. 7. Relation between log P_1 and 1/T in $BaTiO_3$ ceramics

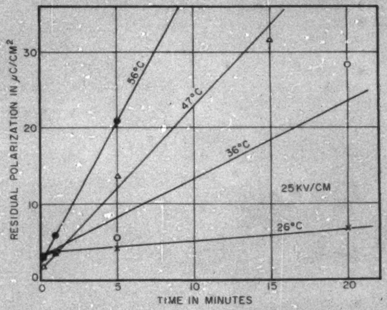


Fig. 8. Relation between residual polarization and time-of-voltage impression in (BaSr)TiO_3 at various temperatures

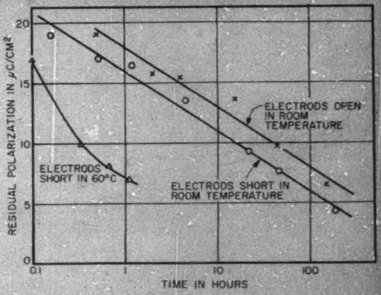


Fig. 9. Decay characteristics of anomalous polarization in $BaTiO_3$ ceramics

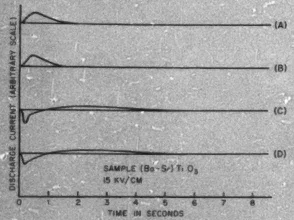


Fig. 10. Oscillograms of discharge current with the time in (Ba-Sr)TiO_3

A—Voltage impression of 1 minute at 15 C
B—Voltage impression of 5 minutes at 15 C
C—Voltage impression of 5 minutes at 60 C
D—Voltage impression of 20 minutes at 60 C

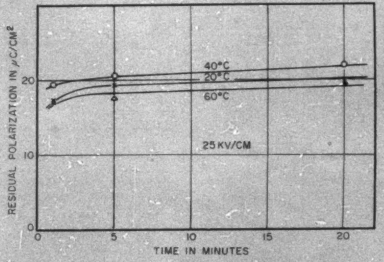
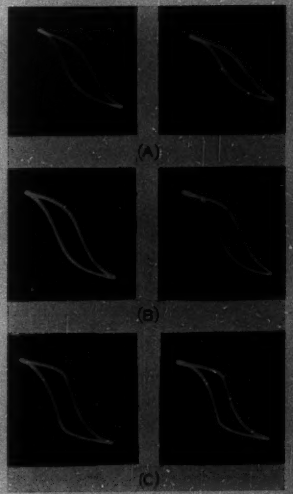


Fig. 11. Relation between residual polarization and time-of-voltage impression in single crystal of $BaTiO_3$

Fig. 12 (right). Hysteresis loops of $BaTiO_3$ ceramic and single crystal before polarization (left) and after polarization (right)

A—$BaTiO_3$ ceramic (polarization: 80 C 20 kv/cm, 15 minutes)
B—$BaTiO_3$ ceramic (polarization: 80 C 10 kv/cm, 30 minutes)
C—$BaTiO_3$ single crystal (polarization: 80 C, 20 kv/cm, 15 minutes)



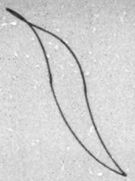


Fig. 13. Unsymmetry of hysteresis loop

point of $BaTiO_3$ is too high and inconvenient for the purpose of our experiment $(Ba\text{-}Sr)TiO_3$ ceramic is used which has a transformation point at 41 C. Fig. 8 shows the result of the experiment. Two straight lines obtained above the Curie point intersect the other two obtained below the point. This naturally is expected because of the nature of P_r which contains P_s below the Curie point.

DECAY OF ANOMALOUS POLARIZATION

Fig. 9 shows the decrease of residual polarization P_r in $BaTiO_3$ ceramics with the passing of time. After being polarized by impressing an electric field of 25 kv/cm for 5 minutes at 60 C, the sample is left at room temperature or at 60 C. Its residual polarization is then measured by the ballistic galvanometer at some selected time. As P_s may also change after some time it is impossible to show the characteristics of P_i accurately. However as the change of P_i is supposed to be comparatively small, as the measurement of the coupling factor shows, Fig. 9 is considered to show the decay of P_i approximately.

The regularity shown in Fig. 7 and the low rate of decay of P_r in Fig. 9 suggests that P_i may be caused by ions emigrating in the ceramics with small mobility.

INTERNAL ELECTRIC FIELD DUE TO ANOMALOUS POLARIZATION

A very interesting phenomenon is observed when P_r is measured in a sample polarized at a temperature above the Curie point. The polarized sample, which is connected to the measuring circuit beforehand, is cooled below the Curie point and then thrown into a paraffin oil bath heated to 170 C. An unusual deflection of the ballistic galvanometer is witnessed, it first deflects instantaneously a bit in the reverse direction and then in the normal direction. In order to verify the phenomenon the development of discharge in polarized ceramics is observed with the electromagnetic oscillograph. Fig. 10 shows the change of discharge current with time. Figs. 10(A) and (B) show the changes corresponding to the samples which have been polarized at 15 C by an electric field of 15 kv/cm and then dipped in a paraffin oil bath heated to 170 C. In the cases of Figs. 10(C) and (D), the samples are at first polarized at 60 C, at a temperature above the Curie point 41 C, then cooled to room temperature, and finally dipped in the heated oil bath. Evidently in Figs. 10(C) and (D) reverse currents are observed at the start of the discharge. Such a phenomenon can be interpreted as follows.

As a sample polarized at a temperature above the transformation point does not have spontaneous polarization P_s but only anomalous polarization P_i, an internal electric field due to P_i will predominate after the removal of the impressed field. This internal field may cause spontaneous polarization when the sample is cooled below the Curie point. The direction of such spontaneous polarization must be opposite to the initial polarization caused by the impressed field. Therefore, if the sample is heated above the Curie point again P_s may be depolarized at first rapidly and then P_i more slowly. The currents induced by them must be opposite in direction to each other. Thus the development of the discharge current should be observed in such a case as is shown in Figs. 10(C) or (D). Moreover the magnitude of P_s caused by the internal field can be evaluated from the deflection of a ballistic galvanometer. The value of P_s thus determined is about 1.8 $\mu c/cm^2$.

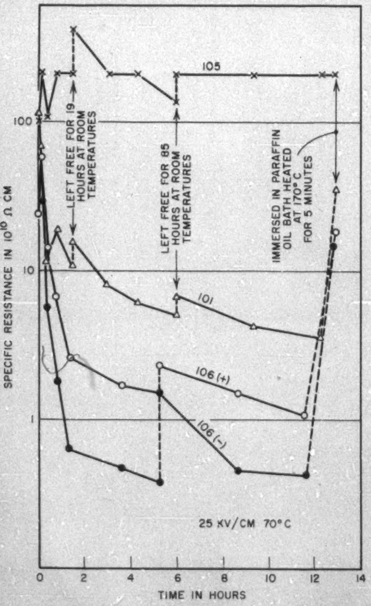


Fig. 14 (left). Change of specific resistance with time-of-voltage impression in $BaTiO_3$ ceramics

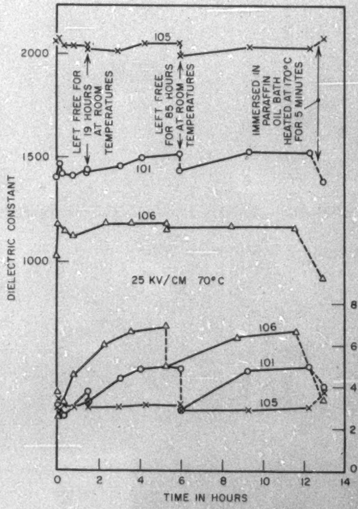


Fig. 15 (right). Change of dielectric properties with time-of-voltage impression in $BaTiO_3$ ceramics

Accordingly it is imagined that the anomalous polarization has a considerable effect on the spontaneous polarization in this way.

RESIDUAL POLARIZATION IN A SINGLE CRYSTAL OF $BaTiO_3$

Fig. 11 shows the relation between the time of impression of the electric field and the result, residual polarization (P_r) of a single crystal of $BaTiO_3$. The crystal is grown in a salt bath having potassium fluoride. Anomalous polarization does not seem very large in this case.

HYSTERESIS LOOP OF POLARIZED SAMPLES

Fig. 12 shows hysteresis loops of polarized $BaTiO_3$ ceramics and a single crystal under various conditions. Sawyer-Tower's circuit is employed in the measurement. The impressed electric field is 20-kv/cm at 50 cycles per second. Although a polarized ceramic shows an unsymmetrical hysteresis loop (Fig. 13), a single crystal seems to keep a symmetrical loop at all times. Such unsymmetry as shown in Figs. 12(A) and (B) is thought to be ascribed to the internal electric field caused by the anomalous polarization as already explained in the previous paragraph.

CHANGE OF RESISTIVITY AND OTHER DIELECTRIC PROPERTIES CAUSED BY ANOMALOUS POLARIZATION

No. *101*, no. *105*, and no. *106* are the samples which have different grades of anomalous polarization from one another. The change of resistivity and other dielectric properties versus time-of-voltage impression are measured for these samples and the results are obtained as shown in Figs. 14 and 15. Although no. *105*, having comparatively small anomaly, shows almost no change in these properties, no. *101* and no. *108*, having large anomaly, show considerable change in resistivity and tan δ. If these samples are left free at room temperature their properties are recovered to some extent as shown by short vertical dotted lines in Figs. 14 and 15. If they are dipped in an oil bath at 170 C for 5 minutes they are recovered nearly to the initial state. As one can see from the result of no. *106* in Fig. 14, resistivity depends more on the polarity of the measuring voltage as P_r grows larger. In these figures a curve having a plus sign shows the resistance measured by applying potential in the same direction as polarizing potential, and another curve having a minus sign is the reversed case. Such different features are also caused by the internal electric field which is built up by the anomalous polarization.

CHANGE OF THE COLOR OF SAMPLES ACCOMPANIED BY ANOMALOUS POLARIZATION

One of the most interesting phenomena accompanying anomalous polarization is the change in the color of the samples. Those samples which show large anomalous polarization change their color from yellowish white to brownish black gradually if they are impressed with a considerably high voltage. But those which do not show anomaly do not change their color at all. If a throughly colored sample is broken in pieces it is found that the interior of the sample has been uniformly tarnished. Even though such a sample is left as it is for a fortnight, the tarnish hardly changes. But it is faded thoroughly when the sample is heated at 170 C for about 10 minutes. Such change of color is thought to indicate some chemical change which has occurred in the sample. This color change resembles that which accompanies the reduction of titanium oxide. Hence anomalous polarization may be caused by some chemical change.

Considerations

Judging from the velocity of its birth and death and also from the temperature characteristics of anomalous polarization, it is almost certain that the phenomenon is caused by the movement of ions. Although it cannot be explained by the action of ions moving along internal surface or grain boundary in ceramics, it may be attributed to some internal defect of the crystal. The phenomenon is not affected by the porosity and grain size of crystals in ceramics. On the other hand, the change of color in a sample accompanied by anomalous polarization proves the occurrence of some chemical change in the crystal. From these facts it might be inferred that the anomalous polarization could be caused by the reduction of Ti in the sample or by the effect of the increased defects of oxygen. A phenomenon resembling the one described was found earlier in TiO_2 by Tago and Watanabe.[4] They investigated the cause of the decrease of insulation resistance and the increase of dielectric loss in a TiO_2 condenser by impressing an electric field and concluded that these phenomena should be attributed to the reduction due to the impressed field which would remove oxygen ions at lattice points in the crystal. Breckenridge and Hosler[5] examined the mechanism of electric conduction in TiO_2 which was partially reduced in hydrogen atmosphere and inferred that some Ti ions were reduced from four to three in valence and, accordingly defects of oxygens should be the result.

Although Blood and his co-workers[3] found an anomalous property of $BaTiO_3$ they attributed it to the state of surface of the crystal to which Könzig and others had alluded. (Könzig[6,7] was of the opinion that the state of the surface layer of $BaTiO_3$ crystal, about 100 Angstroms deep, differs from the interior. The crystal is heavily strained in the layer up to the temperature several hundred degrees centigrade above the Curie point and the strain is caused by the electric field due to the space charge in the surface layer.) But if attention is focused on the considerably large reverse electromechanical activity observed by Blood and others, and also the large inverted spontaneous polarization caused by the internal electric field observed by the authors, it would be more appropriate to attribute the anomaly to the movement of ions through the crystal instead of to the state of the surface layer.

It is very interesting that the anomaly depends to a great degree on the mixing ratio of $BaCO_3$ and TiO_2. When the ratio differs from unity, Ba_2TiO_4, $BaTi_2O_5$, etc., in addition to $BaTiO_3$ might be produced as reported by Kubo,[8] Roy,[9] and others. However, these compounds are all nonferroelectric material. Therefore, if consideration is given to the relation between the anomalous polarization and the spontaneous polarization mentioned in the preceding section, such products are hardly thought much of as the main root of the anomalous polarization. It is more appropriate to consider that the excess or insufficiency of Ti or Ba has a marked influence on the stability of a crystal.

Conclusions

A series of experiments executed on the abnormally large residual polarization in barium titanate is not interpreted by proper spontaneous polarization. It is inferred that the anomaly may be attributed to the movement of vacant holes of oxygen caused by the reduction of Ti in the crystal. It is also explained that the anomalous polarization is remarkably affected by the mixing ratio of $BaCO_3$ and TiO_2.

As anomalous polarization has direct influence on spontaneous polarization and as it causes the decrease of insulation resistance and the increase of tan δ, it should be considered when the ferroelectric substance is to be managed.

References

1. ON THE RESIDUAL POLARIZATION OF BARIUM-TITANATE CERAMICS, **Y. Saito, S. Yamanaka.** *Journal*, Institute of Electrical Engineers (Japan), Tokyo, Japan, vol. 76, 1956, pp. 782–86.
2. ANOMALOUS POLARIZATION IN UNDILUTED CERAMIC $BaTiO_3$, **H. L. Blood, S. Levine, N. H. Roberts.** *Journal of Applied Physics*, New York, N. Y., vol. 27, 1956, pp. 660–61.
3. A MICROSTRUCTURE STUDY OF BARIUM TITANATE CERAMICS, **F. Kulcsar.** *Journal*, American Ceramic Society, Columbus, Ohio, vol. 39, 1956, pp. 13–17.
4. DETERIORATION OF CERAMIC CAPACITORS, **T. Tago, S. Watanabe.** *Journal*, Electrical Communication Laboratory, Tokyo, Japan, vol. 5, 1952, pp. 192–95.
5. ELECTRICAL PROPERTIES OF TITANIUM DIOXIDE SEMICONDUCTORS, **R. G. Breckenridge, W. R. Hosler.** *Physical Review*, New York, N. Y., vol. 91, 1953, pp. 793–802.
6. DAS VERHALTEN VON KOLLIDALEN SEIGNETTEELEKTRIKA III, BARIUMTITANAT $BaTiO_3$, **M. Anliker, H. R. Brugger, W. Känzig.** *Helvetica Physica Acta*, Basel, Switzerland, vol. 27, 1954, pp. 99–124.
7. SPACE CHARGE LAYER NEAR THE SURFACE OF A FERRO-ELECTRIC, **W. Kaenzig.** *Physical Review*, vol. 98, 1955, pp. 549–50.
8. REACTION BETWEEN $BaCO_3$ AND TiO_2 IN SOLID STATE, **T. Kubo, K. Sinriki.** *Journal*, Chemical Society of Japan, Tokyo, Japan, Industrial Chemistry Section, vol. 55, 1952, pp. 49–51.
9. PHASE EQUILIBRIA IN THE SYSTEM $BaO\text{-}TiO_2$, **D. E. Rase, R. Roy.** *Journal*, American Ceramic Society, vol. 38, 1955, pp. 102–13.

Factors Leading to the Development of Ceramic Hydrogen Thyratrons

H. N. PRICE
MEMBER AIEE

A. W. COOLIDGE
MEMBER AIEE

HYDROGEN THYRATRONS are used as switching tubes in line-type pulsers in most high-power radar equipment.[1,2] A lumped-constant transmission line is repetitively charged to a high voltage and discharged by firing the hydrogen thyratron to provide pulses of power to the radio-frequency oscillator tube. The transmission time of the resulting high-frequency pulses to the object scanned and back to the receiver is the mechanism by which distances are measured by the radar equipment.

Development work was started on the hydrogen thyratron in 1941. The use of hydrogen as a gas filling presents several advantages for this service:

1. A high arc drop or cathode fall of potential at high peak current can occur without injury to the oxide cathode. Approximately 600 volts can be tolerated in hydrogen, compared with 30 volts in mercury vapor.
2. A short deionization time is attained in hydrogen, permitting high pulse-repetition rates.
3. It has a high dielectric strength at low pressures.

A line of glass tubes now exists, consisting of seven basic sizes covering the range of peak and average power levels: from 10 kw peak and 25 watts average power to 33 mw (megawatts) peak and 33 kw average power. The essential ratings of these tubes are shown in Table I and the general appearance of a typical glass hydrogen thyratron is shown in Fig. 1.

The trend in tube requirements is toward higher peak and particularly toward higher average-power levels. The necessity for higher power radar equipment is occasioned by the need for detection and measurement at increasingly greater distances in view of the advances in speed of manned aircraft and the advent of guided missiles and nuclear weapons.

Switch tube development has not kept pace with the advancement in the high-frequency oscillator tubes. Former designs in hydrogen thyratrons do not readily lend themselves to great extension beyond present power levels. Since switching tubes of power capacity equal to that of the high-frequency tube are required in each application, a very great need for a basically new design in this field has arisen. The ceramic hydrogen thyratron, with external anode and grid elements, promises to fulfill this need.

Application Background

The elementary circuit of a radar modulator is shown in Fig. 2. The pulse-forming network is charged from the d-c power supply through the charging reactor. It is discharged through the load when the thyratron is triggered.

The charging reactor is a nondissipative element and permits the pulse-forming-network voltage to reach a value in excess of the power-supply voltage. Efficiencies approaching 100% may be obtained in a hydrogen-thyratron modulator using this circuit. The charging reactor and the total capacitance of the pulse-forming network combine to form a resonant circuit. The values of these elements are so chosen that their resonant frequency is usually, but not necessarily, equal to one half the repetition rate at which the modulator is designed to operate. Under such conditions, resonant charging prevails, and current and voltage relationships in the circuit exist as shown in Fig. 3. Charging current appears a series of positive half-cycles of a sine wave while the thyratron voltage appears as a series of half-cycles of a cosine wave swinging about the d-c power-supply voltage as an axis.

When the thyratron is triggered, the shape and amplitude of the discharge current are determined by the circuit constants of the network and load. A theoretical pulse-current wave shape is shown on an expanded scale in Fig. 4. The pulse width is usually set between the extremes of a few tenths of a μsec (microsecond) to several μsec.

The peak inverse voltage existing after the pulse-forming network is discharged aids in deionizing the thyratron, and its amplitude, assuming there is no clipping of the inverse voltage by means of a diode, depends upon the relative impedances of the pulse-forming network and the load. Equations illustrating this and other circuit relationships are given in the Appendix.

Internal Heat Dissipation

The internal heat dissipation in conventional thyratron tubes in rectifier or control service occurs mainly as power delivered to the gas and is equal to the average current multiplied by the tube drop. In the case of hydrogen thyratrons, additional power dissipations which become the ruling factors in tube design occur as a result of the rapid rise and fall of anode current and the high voltages appearing at the initiation and at the cessation of the pulse current.

Paper 58-188, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 31, 1957; made available for printing October 22, 1958.

H. N. PRICE and A. W. COOLIDGE are with the General Electric Company, Schenectady, N. Y.

The major portion of this work was supported by the U.S. Air Force under contract no. *AF19(122)-458*, subcontract *64*, and by the U.S. Signal Corps under contracts no. *DA 36-039-SC-71230* and *DA 36-039-SC-74850*.

Vol.
Pt. 1
1959

Table I. Published Ratings for Existing Glass Hydrogen Thyratron

	1258	3C45	4C35	5C22	5949	5948	1257
Peak forward voltage (kv)	1	3	8	16	25	25	33
Peak current (amp)	20	35	90	325	500	1000	2000
Average current (amp)	0.05	0.045	0.1	0.2	0.5	1	2.6
Peak power output (kw)	10	52	360	2.6×10^4	6.3×10^3	12.5×10^3	33×10^3
Average power output (kw)	0.025	0.07	0.4	1.6	6.3	12.5	33*
Anode dissipation factor (peak voltage × peak current × pulse repetition rate)	0.1×10^9	0.3×10^9	2×10^9	3.2×10^9	6.3×10^9	9×10^9	20×10^9

* At rated peak power.

ANODE DISSIPATION

The voltage, current, and power relations at the beginning and end of the pulse are shown on a greatly expanded time scale in Fig. 5. A peak of power dissipation occurs at the beginning and at the end of each conduction period. Although the time duration of these power pulses is very short (milliμsec) their instantaneous peak is in the mw range and they are repeated hundreds of times per second. The peak anode voltage, peak anode current, and pulse-repetition rate are the major circuit factors which determine anode heating, and their product is the tube rating which appears as the anode dissipation factor in Tables I and II.

At the beginning of the pulse, the rise of current is so rapid, in the order of 10,000 amp (amperes) per μsec, that relatively high anode voltage is required to establish sufficient ionization to support the required current. Consequently, the fall of anode potential from peak forward voltage to arc-drop voltage requires an appreciable time. The resulting power dissipation is very high and, in fact, is the major contribution to anode heating.

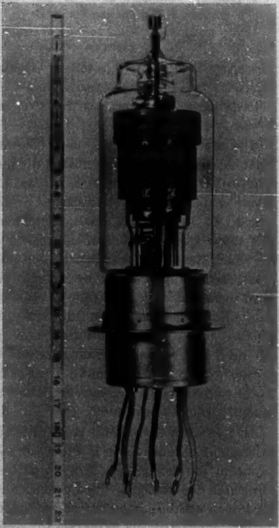


Fig. 1. GL-5948 hydrogen thyratron

At the end of the pulse, an appreciable ion current flows to the anode from the plasma after conduction in the forward direction ceases and inverse voltage is applied. Currents in the order of 50 to 100 amp have been observed in large tubes. An inverse voltage as high as 10 kv appears at the anode at the same time, resulting in a pulse of power dissipation in the tube.

In addition to these peaks of dissipation, which may be considered as the primary sources of anode heating, there is a secondary source which may assume appreciable values. Electrons emitted from the grid may flow to the anode without causing voltage breakdown and gaseous conduction. This is because the distance between anode and grid is in the order of the mean-free path of the gas and the probability of ionization is low. While the resultant currents may be in the milliampere range, the power dissipated can be high since the electron-accelerating voltage is the network forward voltage which becomes as high as 30 to 50 kv during parts of the charging cycle. In addition, the current flows essentially throughout the interpulse interval.

GRID DISSIPATION

The power delivered to the gas in the steady discharge is distributed to the electrodes surrounding the discharge. The greater portion of this is delivered to the grid because of its large exposed area. It is approximately equal to the average current multiplied by the tube drop.

CATHODE DISSIPATION

Because of the extremely high peak currents, 1,000 to 2,500 amp, and correspondingly high rms currents, the dissipation in the cathode results largely from the passage of the emitted electron current through the resistance of the coating. Its principal effect is to heat the cathode, which increases the rate of evaporation of active material and causes depletion of the coating.

General Design Considerations

The hydrogen-thyratron design reflects the unusual operating conditions in pulse switching service:

1. Extremely high peak current and voltage levels.
2. Current pulses in the μsec range with attendant high rates of current rise and fall.
3. High repetition rates with correspondingly short lengths of time available for recovery of the control function of the thyratron.
4. Rapid application of high inverse voltage to the anode in presence of very high levels of residual ionization.

A cross section of a typical high-power hydrogen thyratron of present design is shown in Fig. 6. The tube is approximately 6 inches in diameter and 16 inches long and is rated at 12 mw peak and 12 kw average power. Its envelope is of conventional glass design with tungsten lead seals for supporting and making connections to the internal elements. The anode is a large disk supported from a lead at the top of the tube. Molybdenum is used for the anode to withstand the severe ion bombardment which occurs after

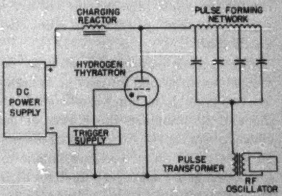


Fig. 2. Schematic diagram of typical modulator circuit

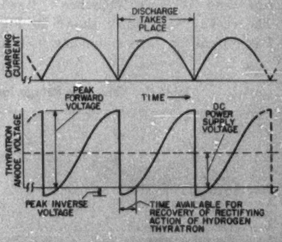


Fig. 3. Current and voltage wave shapes in charging circuit of radar modulator

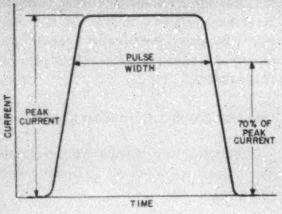


Fig. 4. Idealized output pulse current wave shape for radar modulator

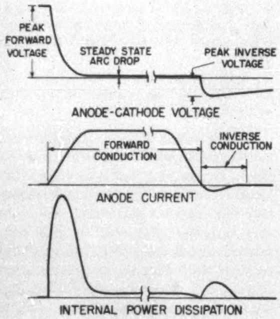


Fig. 5. Tube voltage, current, and wattage dissipation during pulse

conduction when high inverse voltage is rapidly applied to deionize the grid-anode space.

The hydrogen pressure within the envelope is established initially at as high a level as possible to minimize anode heating and to minimize the effects of gas clean-up during the life of the thyratron. In view of the high pressure, anode-to-grid spacings are made very small for high-voltage operation. The voltage breakdown between two electrodes in a gas follows the well-known curves of breakdown voltage versus the product of pressure and distance (Paschen's Law). The curve for mercury shown in Fig. 7 is typical for gases commonly used in tubes. It extends well into the left-hand region for which no values for hydrogen are published. The grid-anode space in the thyratron is such that the operating points are in this region where the breakdown voltage is rising rapidly as the product of pressure and spacing is reduced. At a given pressure, the voltage at which breakdown will occur depends upon the longest path between grid and anode over which a discharge will take place.

In view of this, the small distance from grid to anode must be maintained throughout and the grid structure is designed to surround the anode completely. Long-path discharges at the anode stem are prevented by surrounding the stem with a glass cylinder which fits into a metal sleeve in the grid structure. The anode structure becomes more complex in the larger tubes which operate at higher voltages and at higher temperatures. Additional members, as illustrated in Fig. 6, are required to reduce voltage gradients and to avoid conduction through the glass at high temperature. Such conduction might complete the electrical path necessary to form a long-path discharge.

Hydrogen thyratrons are designed with a positive control characteristic and differ from most thyratrons in this respect. This simplifies the grid-trigger circuit in that no grid bias is required. In addition, the amount of baffling associated with the positive characteristic minimizes grid-emission difficulties. In order to achieve these characteristics, the baffling between cathode and grid shown in Fig. 6 is introduced. The grid itself is a perforated or slotted disk and the baffles underneath it are solid.

The cathode is a conventional, oxide-coated, indirectly heated type with sufficient area to furnish the peak emission and to conduct the required rms current through the coating without raising its temperature beyond the permissible limit. This limit is a function of its evaporation rate versus temperature and the expected life.

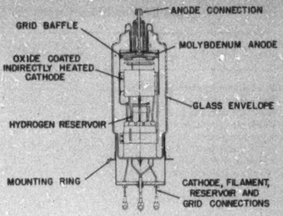


Fig. 6. Cross-sectional view of GL-5948 hydrogen thyratron

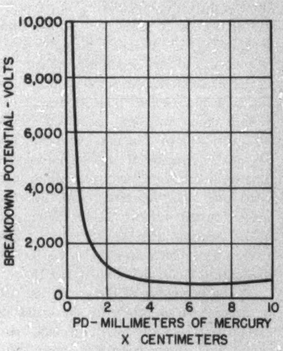


Fig. 7. Breakdown voltage between two electrodes in mercury vapor as a function of the product (Paschen's Law) of the vapor pressure P and the gap length D

Table II. Objective Ratings for New Ceramic Hydrogen Thyratrons

	Z-5069	Phase I Z-5101	Phase II Z-5212	Phase III
Peak forward voltage (kv)	22	40	50	100
Peak current (amp)	2000	2400	2000	2000
Average current (amp)	4	2.2	8	8
Peak power output (kw)	33×10^3	48×10^3	50×10^3	100×10^3
Average power output (kw)	66	44	200	400
Anode dissipation factor (peak voltage × peak current × pulse repetition rate)	40×10^9	45×10^9	400×10^9	800×10^9

The disappearance of gas from an enclosed container in which an electrical discharge is present is familiarly known as clean-up. It is recognized that various processes contribute to the over-all result.

In noble gases, it has been shown[4] that clean-up occurs as a result of the impact of positive ions upon electrodes at considerable negative potentials with respect to the source of ions. Sometimes the noble gas ions stick in the electrode to which they have been accelerated. At other times, apparently, they are neutralized there and leave with enough velocity to cause them to stick on adjacent surfaces. Sputtering of electrode metal on to other surfaces may enhance the clean-up by continually providing fresh surfaces and also by mechanically carrying and trapping the gas.

In hydrogen thyratrons these mechanisms undoubtedly occur, but they are not alone in producing the effect. The hydrogen is dissociated in the arc to atomic hydrogen which is extremely active chemically. This may be expected to cause clean-up by chemical action[5] or by diffu-

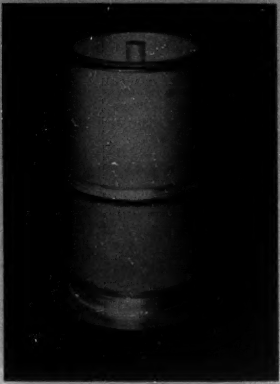


Fig. 8. Developmental ceramic hydrogen thyratron Z-5069

sion through certain materials forming the boundaries of the arc.

The mechanisms of gas clean-up become sufficiently severe in high-power hydrogen thyratrons that a hydrogen replenisher or pressure-maintaining "reservoir" has been developed for these tubes. It consists of a quantity of titanium or zirconium hydride held at an elevated temperature by an auxiliary filament. The hydride dissociates according to the reversible reaction:

$$T_iH_2 \rightleftarrows T_i + H_2$$

The equation moves to the right or left depending upon the temperature of the hydride, the concentration of the hydrogen remaining in the hydride, and the pressure of hydrogen above the hydride. A hydrogen reservoir provides a means for deriving the optimum pressure for tube operation and is a replenishing source for minimizing the effects of hydrogen loss.

With the reservoir characteristics known, the pressure range over which the tube will operate successfully can be determined by varying the voltage applied to the reservoir filament. The voltage range over which successful operation is attained is usually referred to as reservoir range. With a given reservoir characteristic, its width provides a measure of the performance margin at a given level of operation. The range is most accurately presented as a percentage of the center or optimum operating point.

The upper end of the operating pressure range and hence the upper end of the reservoir range is a function of voltage breakdown and of recovery time. As previously indicated, thyratrons operate to the left of the minimum point of the curve shown in Fig. 7 and, with a given electrode spacing the voltage, breakdown occurs at progressively decreasing levels with increasing pressure.

Either of several criteria may determine the lower end of the operating pressure range. The amount of ionization from the trigger pulse required to initiated conduction to the anode within the specified anode firing time and within the specified variation in successive firing times, "jitter," increases with decreasing pressure. Pressures below an approximate value of 200 microns make it impossible for anode conduction to take place at all.

The establishment of sufficient plasma density to support the required rate-of-rise in anode current and the ultimate peak value of anode current becomes increasingly difficult with decreasing pressure. The voltage required to establish and maintain the plasma rises accordingly and anode dissipation rises rapidly. Excessive anode heating is the most common factor limiting the lower end of the pressure range.

Ceramic Hydrogen Thyratrons

The ceramic hydrogen-thyratron design provides greater reservoir ranges at present power levels and acceptable ranges at greatly extended power levels. The *Z-5069* developmental tube is shown in Fig. 8. The tube is approximately 5 inches in diameter and 10 inches tall. The objective ratings are 33 mw peak and 66 kw average power. The internal construction is shown in the cross-sectional drawing, Fig. 9. Both anode and grid consist of cups with extended flanges brazed directly to the ceramic cylinders which form the envelope and provide the insulation between elements. The cathode is supported from a header brazed into the bottom of the ceramic envelope.

Extension of the lower end of the reservoir range at high average-power levels is accomplished in several ways. Because it is external to the tube and is convection rather than radiation cooled, the heat-dissipation capacity of the anode is an order of magnitude greater than in internal anode designs. In addition, the design lends itself very readily to forced liquid cooling for still greater heat-dissipation rates.

The reduction in grid temperature is twofold. More cooling is provided by the external flange and its connection to the envelope, and less heat is radiated to it from the relatively cool anode. The dissipation at the anode is further reduced by the reduction in electron current emitted to it from the grid because of the lower grid temperature.

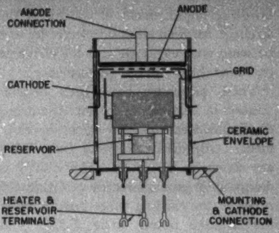


Fig. 9. Cross-sectional view of developmental ceramic hydrogen thyratron Z-5069

Extensions in the upper end of the pressure range affect the average power level mainly by raising the permissible anode voltage. These occur chiefly because of improved high-voltage operation afforded by the simple planar construction in the anode-grid area. The spacings between the cylindrical surface of the anode and grid and the corresponding ceramic surfaces are less than the mean-free path of the gas. Electrons in this space are much more likely to be accelerated to a near-by surface than to cause ionizing collisions with gas molecules. Thus, the probability of ion formation in this area is extremely low and subsequent long-path discharges are prevented. Voltage-breakdown levels between grid and anode near the theoretical value for the spacing between them are thus achieved.

The ceramics are dimensionally very accurate since they are ground to size. The method of assembly, wherein the elements are stacked and brazed together, preserves and utilizes this accuracy. The

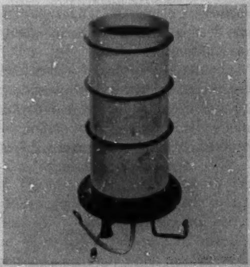


Fig. 10. Developmental ceramic hydrogen thyratron having one gradient grid, Z-5101

resulting improvement in control of the spacings between adjacent electrodes and between each electrode and the envelope enhances high-voltage performance.

These principles have been demonstrated in ceramic hydrogen thyratrons of extended ratings now under development. Objective ratings for these tubes are shown in Table II. The *Z-5069* is designed for peak power levels within the range of the existing tubes but for twice the average power. The *Z-5101* and *Z-5212* represent the first two phases in a co-ordinated program to provide progressively larger sizes in terms of both peak and average power. The developments of the *Z-5069* and the *Z-5101* are substantially complete. The development of the *Z-5212* will, when completed, form the basis for extension to the third or final phase of the current program.

Successful operation and acceptable reservoir ranges have been achieved at the objective power levels. Life tests are being conducted to the limit of existing test equipment. Paralleling of the largest available test equipments in pairs will provide sufficient power on a continuous basis for extensive life-test evaluation at full ratings.

The *Z-5101* has been successfully operated at peak anode voltages to 50 kv while switching a peak-output power level of 75 mw. Performance at this and higher voltage levels, as in phases II and III of the current development program, is facilitated by the use of voltage-dividing or gradient grids. The anode voltage is divided between the tube sections thus formed by means of external circuitry or simply by the internal capacitance between the elements. The voltage between adjacent electrodes is thereby reduced to values at which adequate safety margins are attained. A *Z-5101* with one voltage-dividing grid (two sections) in addition to the usual control grid is shown in Fig. 10. The stacked or building-block type of structure in the ceramic-envelope design lends itself admirably to the inclusion of these additional elements.

Comparatively high mechanical strength and rigidity and correspondingly improved tolerance for shock and vibration are also features of this construction. These are especially important in view of the almost exclusive application of these tubes in military service.

PROBLEMS IN FURTHER EXTENSION

As the tube ratings increase, anode dissipation becomes a greater problem. Great increases in anode size have doubtful value in view of the probability of decreased utilization of the total area of both anode and grid. Consequently, further increases in power levels to phases II and III will probably required forced liquid cooling of the anode and possibly the grid.

Hydrogen clean-up becomes increasingly severe as power levels are raised. A solution of the basic problem must be found or the use of larger reservoirs or maintenance of tube pressure by a regulating means employing a reservoir and a monitoring device must be employed.

With forced cooling the size of the envelope may be determined by the cathode area rather than by the anode-and-grid structure. New types of cathode coatings are being evaluated to provide sufficient emission in smaller sizes. These coatings will have lower internal resistance than the conventional oxide coating and, therefore, there will be less heating by conduction losses.

Conclusion

Performance tests indicate that the ceramic, external anode-and-grid design permits extension of hydrogen thyratron ratings to the desired objectives. Thus, ceramic hydrogen thyratrons will provide the required switching capacity for contemplated higher-power radar equipment.

Appendix

Simple relations are given for the following: 1. inverse voltage, 2. forward voltage, 3. peak current. 4. average current. 5. peak power., and 6. average power.

Some of the following equations are approximations and depend on these assumptions: 1. There is no clipping of the inverse voltage by means of a diode. 2. There are no losses in the charging circuit or in the hydrogen thyratron.

Inverse voltage:

$$e_{px}=e_{py}\frac{Z_0-R_L}{Z_0+R_L}$$

where

e_{px} = peak inverse voltage
e_{py} = peak forward voltage
Z_0 = characteristic impedance of pulse forming network
R_L = load impedance reflected at the primary of the pulse transformer

Forward voltage:

$$e_{py}=2E_{dc}+e_{px}=E_{dc}\frac{Z_0+R_L}{R_L}$$

where

E_{dc} = d-c power supply voltage

Peak current:

$$i_p=\frac{e_{py}}{Z_0+R_L}$$

where

i_p = peak anode current

Average current:

$$I_p=i_p\times t_p\times p_{rr}$$

where

I_p = average anode current
t_p = pulse width
p_{rr} = pulse repetition rate

Peak power:

$$P_{pk}=E_{dc}\times i_p$$

where

P_{pk} = peak power delivered by the modulator to its load

Average power:

$$P_{avg}=E_{dc}\times I_p$$

where

P_{avg} = average power delivered by the modulator to its load

References

1. PULSE GENERATORS (book), G. N. Glasoe, J. V. Lebacqz. McGraw-Hill Book Company, Inc., New York, N. Y., London, England, 1948, pp. 335–54.

2. HYDROGEN THYRATRONS, H. Heins. *Electronics*, New York, N. Y., vol. 19, July 1946, pp. 96–102.

3. Glasoe and Lebacqz, *op. cit.*, pp. 175–496.

4. CLEAN-UP OF HELIUM GAS IN AN ARC DISCHARGE, M. J. Reddan, G. F. Rouse. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1924–29.

5. SCIENTIFIC FOUNDATIONS OF VACUUM TECHNIQUE (book), S. Dushman. John Wiley & Sons, Inc., New York, N. Y., 1949, pp. 654–87.

Vol.
Pt. 1
1959

Electromechanical Impedance: Analogs and Duality

WAYNE B. SWIFT
ASSOCIATE MEMBER AIEE

IT IS FELT that 4-terminal network theory is potentially of great value in the analysis of problems in the non-electrical realm and problems which are only partially electrical. In attempting to apply 2-port theory however,[1,2] some perplexing difficulties arise. These appear to the author basically to be due to somewhat loose habits of thought in defining impedance and admittance on the basis of purely electrical concepts without due regard for their significance in other domains.

The purpose of this discussion is to present a re-examination of the concept of impedance and to suggest a consistent criterion for impedance character. This requires careful specification of network variables. It is found on topological grounds that there is a natural correspondence between current in an electric system and force in a mechanical system. The use of this fact allows us to determine whether a given lossless device is properly considered as an ideal transformer or as a gyrator and whether a given parameter is properly considered as an impedance, an admittance, or as an h-parameter.

Finally, it is pointed out that the curious differences between the all-electric equivalent-circuit behavior for magnetic transducers and electrostatic transducers appear because one is, properly speaking, an ideal transformer and the other is, properly speaking, an ideal gyrator.

Definition of Network Variables

A self-immittance is a ratio between two variables of a network element which are so chosen that at least the following two conditions apply: 1. their product is power; and 2. either is convenient for use in establishing the contribution of the element in question toward system equilibrium. The second of these conditions suggests that an equilibrium principle should pertain to each of the two variables so used. Equilibrium can always be guaranteed in a network configuration by writing equilibrium equations which result from a systematic choice of one sort of variable in terms of the other. In an electric network, for example, mesh current equations can be written to express voltage equilibrium, the current equilibrium conditions being automatically satisfied because of the systematic choice of the currents. These equations represent a condition of equilibrium about a topologically doubly (at least) connected region, i.e., about a mesh of the network.

No amount of information at any particular point can possibly suffice to give all the information which is needed to state the equilibrium conditions around any mesh. For this it is necessary to imagine a complete traverse of a boundary of a multiply connected region, something which, by definition, is not reducible to a point. A formally similar equation can be written, however, if we first obtain the dual of the original network (this presumes that the original network was planar). Then a node-voltage equation of identical form is written by establishing current equilibrium at a point. The equation asserts the existence of this equilibrium at the point which is the node.

This discussion has suggested that a third condition on the variables to be used is that they must be of the nature of duals of one another, and this means that one must achieve equilibrium about a mesh while the other achieves equilibrium at a point.

The conditions on the two variables used to define immittances may then be restated: 1. one (e.g., m) obeys an equilibrium condition about the mesh; 2. the other (e.g., n) obeys an equilibrium condition at the node; and 3. the instantaneous product mn yields power into the element being considered.

The value of either m or n in an element is known from the properties of the element whenever the value of the other variable is known.

It is customary to express this fact in the form of defining relations of the form of equation 1

$$Z=\frac{m}{n} \quad Y=\frac{n}{m} \tag{1}$$

where all symbols are considered as in general operational quantities.

Mechanical Networks and Electrical Analogs

In electric network analysis, it is clear that the m variable is voltage while the n variable is current. But in the electrical domain there has never been any particular confusion over the question of whether a given ratio is an impedance or an admittance; it has appeared to be obvious. In the mechanical domain, however, it had been taken for granted for a long time that force was a variable which directly corresponded to voltage and all equivalent circuits were based on this assumed correspondence. It was not until relatively recently that it was pointed out by Firestone[3] that other equivalent circuits could be evolved using current and force as corresponding variables. The question of whether or not voltage-force was the only allowable analogy was settled by the discovery of Firestone's "mobility" analog, but only at the cost of raising other and perhaps more serious questions. Of the two analogs available, was either of them a "true" one? If so, which one? Why did the current-force analog preserve the topology of the mechanical system for a magnetic transducer but not for an electrostatic transducer? These questions are fundamental to a consideration of the reciprocity of the systems which are not purely electrical.[4] Using the force-voltage correspondence for a d-c motor, the z-matrix turns out to have opposite signs on the z_{12} and z_{21} elements and is therefore termed antireciprocal. Using the force-current correspondence, however, the z-matrix turns out to be symmetrical. The device has not changed, only the observer's point of view has shifted. Yet, if the self-impedances are neglected, in one case the device is an ideal transformer while in the other it is an ideal gyrator.

Ideal Transformers and Gyrators and Corresponding Variables

Investigations of the z-matrix and the h-matrix of the transducers are perhaps the most suggestive way of attacking these questions since the results tend to parallel those which appear for ideal elements in the purely electrical domain.

Paper 58-529, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Great Lakes District Meeting, East Lansing, Mich., May 5–7, 1958, and re-presented at the Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted March 28, 1957; made available for printing October 22, 1958.

WAYNE B. SWIFT is with The University of Wisconsin, Madison, Wis.

An ideal transformer, for example, has no proper z-matrix. All elements in the matrix are infinite unless external elements, unconnected with the essential mode of operation of the transformer, are included to furnish spurious interdependences between properly independent quantities. If an ideal transformer is attacked in any way such as to suggest that its z-matrix is well behaved, then there must be something odd and improper about the approach used. Its h-matrix is perfectly well behaved, however. An ideal gyrator, on the other hand, is characterized by a well-behaved but antireciprocal z-matrix; its h-matrix contains only infinite components. Thus, when a device called a gyrator is shown to have a normal h-matrix but an ill-behaved z-matrix, it is certain that the approach to the problem was unnatural and somehow improper.

As an example of the sort of improper throught process which can be imagined in this regard, one may suppose that the quantities of primary voltage and secondary current were regarded as corresponding to the m variables. Then the "impedance" matrix could be written as shown:

$$m = zn = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} \qquad (2)$$

Now the elements of this matrix will be as follows, if the self-impedances are ignored:

$z_{11} = 0$

$z_{12} = 1/a$ (turns ratio)

$z_{21} = -1/a$

$z_{22} = 0$ (3)

There are two things seriously wrong with this z-matrix for an ideal transformer: First, it is antireciprocal despite the fact that an ideal transformer is well known to be a reciprocal element, and second, it is well behaved instead of containing only infinite elements as the z-matrix of an ideal transformer should. The conclusion is that the choice of m_1, m_2, n_1, n_2 as e_1, i_2, i_1, e_2, respectively, is improper and unnatural since such a choice leads to elements of the z-matrix which are not dimensionally equivalent to an impedance. This is something which can be seen intuitively when purely electrical problems are considered.

Corresponding Electrical and Mechanical Variables

In the mechanical domain confusion exists since no one can be sure what impedance means. If the variables m and n are selected according to the rules already mentioned, however, the identity of mechanical impedance is at once established. Furthermore, such a procedure gives light on what ought to be done in the process of choosing corresponding variables at opposite ends of a transducer problem. Corresponding variables are those which have corresponding topological equilibrium equations. Thus, force, which obeys the condition of equilibrium at a point, takes on a likeness to current in the electrical realm. Velocity, which must be summed about a mesh of the mechanical system to express equilibrium, is likened to voltage. If this rule is followed, the resulting matrices for ideal elements fall into one of these two categories:

1. If the ideal element is truly an ideal transformer, its z-matrix will be ill behaved, but its h-matrix will be well behaved and antireciprocal.
2. If the ideal element is truly an ideal gyrator, its z-matrix will be well behaved and antireciprocal but its h-matrix will be ill behaved.

Transducers

It can now be pointed out that if the ideal electromagnetic transducer is considered under these rules, its h-matrix is well behaved and the off-diagonal elements are opposite in sign. This indicates a reciprocal element which is an ideal transformer. If the force-voltage choice of variables had been made, a well-behaved z-matrix would have been obtained with the antireciprocal character.

If the z-matrix is written for the ideal electrostatic transducer with proper choice of variables, i.e., with force-current correspondence, there results a well-behaved and antireciprocal matrix. Thus the electrostatic transducer is a gyrator. In order to include representations of the mechanical elements of the electric transducer in an over-all electric-system diagram, it is therefore necessary to include an ideal gyrator. If for any reason the gyrator is to be eliminated from the over-all diagram then it is necessary to take the dual of either the electrical or the mechanical portion of the over-all system (not both). This procedure yields an ideal transformer in the place formerly occupied by the gyrator. The custom has been to obtain the dual of the mechanical part of the system in these situations, and this procedure yields an apparent force-voltage correspondence. The process of obtaining a semidual configuration to rid the system of its gyrator effects is unnecessary for the magnetic transducer since this device contains no ideal gyrator. Hence the force-current analog yields an electrical equivalent of the mechanical part of a magnetic transducer which can be connected into the electrical part of the over-all diagram with no gyrator but still holds the original topology of the mechanical network.

It seems to the author that the better procedure in the case of the electrostatic network is to include the ideal gyrator specifically when indicating the diagram of the mechanical part of an electrostatic transducer in order to 1. maintain the original topology, and also 2. to force recognition that a true gyration has taken place. This is inescapable when the same electric circuit is made to feed two transducers, one electric, the other magnetic, whose mechanical elements are interconnected.

It is now possible to demonstrate the way in which the classical approach to electrical analog problems fails. This is usually not apparent because problems of such simplicity are used. For instance, a mechanical network having five nodes with ten elements linking the five nodes in all possible ways will result in a network that is not planar[5] and hence has no dual. For this reason it is impossible to draw an electrical equivalent circuit using the force-voltage analog unless obviously nonequivalent elements are used. In all such cases as this the mobility analog shows its true character since it can easily be drawn and since it reproduces the topology and the active or passive character of the corresponding elements of any network as well as formally identical equations.

The conclusion is, then, that the classical analog is a false one and that the force-current analog is the one true analog.

Conclusions

In analyzing any network element a pair of related variables must be used such that over-all system equilibrium implies equilibrium at a connection point in one variable n, and about a multiply connected region in the other, m. Impedance is then defined as m/n.

Application of this principle leads to the conclusions that:

1. Force and current are corresponding quantities.
2. A force-voltage analog in a transducer is artificial and requires introduction of an ideal gyrator and semidual topology into the equivalent circuit. This gyrator can

be eliminated only by taking the dual of one portion of the circuit. The whole process is impossible except on planar systems.

3. A transducer using an electric field as the coupling medium is in essence a gyrator; an over-all electrical equivalent circuit contains an ideal gyrator. Elimination of this gyrator requires obtaining the dual of one part of the circuit separated by the gyrator.

4. If a network containing one magnetic and one electrical transducer provides interconnections both electrically and mechanically, then its over-all diagram must contain an ideal gyrator.

The case of a nonplanar mechanical network is examined to test these conclusions and it is shown that an electrical equivalent circuit can readily be found provided force and current are regarded as corresponding quantities. But no electrical equivalent circuit exists if force and voltage are regarded as corresponding quantities, presuming that an equivalent circuit implies element-by-element correspondence. By this test, the validity of the force-current analog and the falsity of the force-voltage analog is confirmed.

References

1. Analysis of D-C Rotating Machines as Two-Port Networks—I, V. C. Rideout. *AIEE CP 57-760* (available on request).

2. Analysis of D-C Rotating Machines as Two-Port Networks—II, W. B. Swift. *AIEE CP 57-761* (available on request).

3. A New Analogy Between Mechanical and Electrical Systems, F. A. Firestone. *Journal*, Acoustical Society of America, New York, N. Y., vol. 4, 1933, pp. 249–67.

4. A General Theory of Passive Linear Electroacoustic Transducers and the Electroacoustic Reciprocity Theorem, L. L. Foldy, H. Primakoff. *Ibid.*, vol. 17, 1945, p. 109; vol. 19, 1947, p. 50.

5. Sur le Probleme des Courbes Gauches en Topologie, S. Kuratowski. *Fundamenta Mathematicae*, vol. 15, 1930, pp. 271–83.

Measurements of Narrow-Band Noise on Telephone Facilities in Connection with Analog Data

J. O. EDSON
NONMEMBER AIEE

F. E. FROEHLICH
ASSOCIATE MEMBER AIEE

R. K. TOWNLEY
NONMEMBER AIEE

THE RECENT GROWTH of automation and the handling of data by machines have resulted in a growing need for data communication between terminals separated by many miles. Both the trend toward decentralization of business and the tendency toward centralized data-processing centers have contributed to this need. One solution to the problem of providing effective data communication is to use the telephone network which makes the interconnection of almost any two points in the country feasible.

Facilities for the transmission of digital data are presently available.[1,2] However, many of the data sources are analog in nature. Also most of the transmission facilities are designed to handle analog signals. Thus, the transmission of analog data directly instead of converting to digital form could lead to more simple terminal equipment and should therefore be investigated.

The telephone network has been designed primarily for the transmission of speech in an acceptable and economical way. The acceptable limits for the various transmission characteristics of the telephone network have been chosen principally to meet this requirement. However, speech signals can tolerate surprisingly large distortions of varied sorts and still preserve acceptable quality. The question then arises: what limitations on the transmission of analog data are imposed by the telephone network? Also what direction should the design of special systems take so as to work in the environment of the speech network.

It is well known that the telephone plant has been adapted satisfactorily to a number of nonvoice uses. A partial list would include telemetering, facsimile, telephoto, radio network, standard frequency transmission, and digital data transmission. Each of these services posed some problems which have been solved satisfactorily.

A high-resolution narrow-band analog-data system, which was designed on the basis of the usually measured characteristics of voice facilities, was tested. The tests were performed over a variety of transmission facilities typical of those used in both the regular switched network and private-line service. Since the co-ordination with supervisory and switching systems was not studied, these results should be interpreted as referring only to private-line services. As a result of these tests it was found that, in some carrier telephone systems, side bands are generated which appear as narrow-band noise around a transmitted tone. While these side bands were sufficiently low in amplitude that they could hardly be detected by the human ear, they seriously affected the performance of the analog-data system under test. The measurements of this noise and ways of overcoming its effects are discussed.

Investigation of Analog-Data Transmission

An interesting example of analog-data transmission that has recently been studied is telewriting, which may be defined as the sending of handwritten messages over a transmission medium. Investigations have shown that satisfactory performance may be attained over three channels, each 10 cps (cycles per second) wide. Differential delay between any two of the three channels should be limited to ±4 milliseconds. Resolution of one part in 250 is required. This resolution implies a signal-to-noise ratio of 60 db (decibels) when provision is made for typical noise peaks encountered in white noise. The requirements of many other analog-data systems can be included within these limits when due consideration is given to the number of channels, bandwidth, and signal-to-noise ratio required.

The simplest method of deriving three data channels from a voice channel is to use double side-band amplitude modulation on three carriers suitably located in the voice band. However, consideration of the permissible signal levels and the expected noise levels indicates that a signal-to-noise ratio as low as 35 db would be

---

Paper 58-1241, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting Pittsburgh, Pa., October 26–31, 1958, and re-presented at the Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted August 19, 1958; made available for printing December 3, 1958.

J. O. Edson, F. E. Froehlich, and R. K. Townley are with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

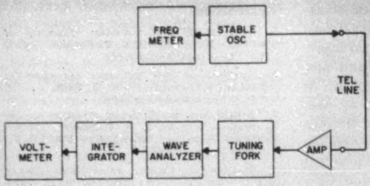


Fig. 1. Block diagram of narrow-band noise measuring system

frequently encountered. Thus, in order to achieve the required 60-db ratio, the use of frequency modulation seems indicated. A frequency-modulated wave, equal in amplitude to a fully amplitude-modulated wave, with a total deviation of ±52 cps will give the required additional 25 db advantage over noise.

A telewriting system which met the preceding requirements, was tested over various facilities. A sample representative of normal telephone transmission facilities was chosen for the tests.

Prediction of performance over each facility was made on the basis of noise measurements with the facility terminated. Readings were taken with a Western Electric Company *2B* noise meter.[3] For the compandored carrier-telephone system,[4] due consideration was given to the changes of gain within the system that are produced in the compandor by the presence of a tone.

In most cases, the performance of the telewriting system was in excellent agreement with the predictions based on normal measurements of the voice facilities. However, the performance was significantly poorer than predicted over types *N*, *O*, and *ON* carrier systems.[5,6] The *N*, *O*, and *ON* carrier systems were designed as inexpensive compandored systems and they make full use of the noise advantage the compandor gives on voice.

The most striking feature of the observed disturbance of the written message when using these carrier systems is its independence of large variations in subcarrier amplitude. This suggests that the disturbing signal maintains a constant relationship to the subcarrier and manifests itself only while the carrier is present. Preliminary tests indicated that the mechanisms involved are apparently frequency insensitive. That is, the noise spectrum about the test signal preserves its shape as the signal is varied over a narrow range.

MEASURING EQUIPMENT

Consideration of the magnitude and location of signals that can cause the observed disturbance leads to the following requirements on the measuring equipment:

1. Selectivity giving approximately a 1-cps bandwidth.
2. Selectivity capable of resolving products 50 to 60 db below a test tone over a band 10 cycles on either side of the test tone.
3. Provision for approximately 100 db dynamic range.
4. Stability over periods sufficiently long to allow a series of measurements to be made without readjustment.
5. Integration of the detected output sufficient to provide a reasonable estimate of the noise.

A block diagram of the test setup is shown in Fig. 1. It consists of a stable oscillator whose frequency can be varied with precision over the band of interest. A tuning fork and wave analyzer provide the required selectivity and sensitivity. The integrator and d-c vacuum-tube voltmeter make a reasonably accurate estimate of the noise reading possible.

A Times Facsimile fork model *FK2B* was used. The relative transmission versus frequency characteristics are shown in Fig. 2. The fork used was resonant at 1,600 cps with an effective bandwidth slightly less than 1 cps. The resonant frequency was found to depend only slightly on signal amplitude and remained stable to a few parts per million. It may be seen from Fig. 2 that the maximum loss is approximately 40 db which, by itself, is insufficient for the proposed measurements.

The selectivity of the General Radio *736A* wave analyzer was used to supplement the fork. A plot of its selectivity is shown on Fig. 3. This curve shows a relatively flat maximum which may lead to a shift of the curve relative to the fork. Tests showed that an allowance of ±05. cycles was sufficient to provide for mistuning and drift. Fig. 4 gives the computed over-all selectivity curves for the extremes of ±0.5 cycles mistuning. Compared with these curves are the results of several calibration runs made during the course of the tests. The comparison indicates that the calibration is stable and will give reliable measurements in the range of interest. Further, graphical integration of the curve gives an effective bandwidth of 0.7 cps for the combination.

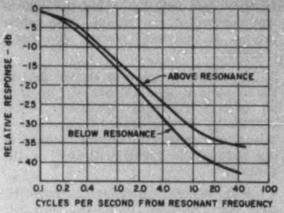


Fig. 2. Relative response versus frequency of tuning fork near resonance

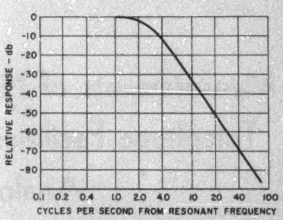


Fig. 3. Relative response versus frequency of wave analyzer

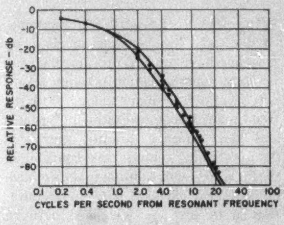


Fig. 4. Over-all response of narrow-band noise detector

The charge time constant of the integrator in Fig. 1 is approximately 10 seconds while the discharge time constant is 500 seconds. Successive readings taken on a white-noise source were reproducible within ±0.5 db when the sample time was 1.5 minutes. This period was used in the measurements.

To calibrate the instrument, a signal consisting of a tone mixed with white noise was applied to the input terminals. The noise was 3-kc band-limited white noise and the signal-to-noise ratio was 20 db. On the basis of an effective bandwidth of 0.7 cps the readings should approach an asymptote 56.3 db below the

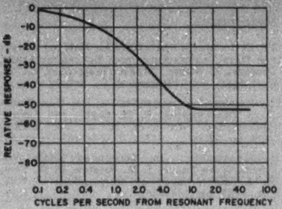


Fig. 5. White-noise calibration of system

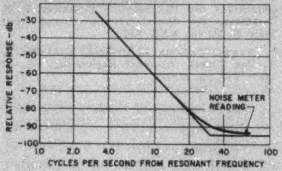


Fig. 6. Noise measurement of cable facility

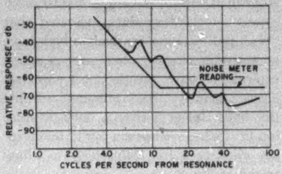


Fig. 7. Noise measurement of type O carrier telephone system

tone. The result of this measurement is plotted on Fig. 5. The final asymptote is 53.5 db below the tone. The slightly higher reading is reasonable in view of the peak reading nature of the indicating circuit.

NARROW-BAND NOISE MEASUREMENTS

The narrow-band noise measurements were made on a looped basis, that is, both the transmitting and receiving ends of the circuit were in the same laboratory. Precautions were taken to insure negligible direct coupling between generator and detector. The amplitude of the received signal was brought to a standard amplitude by a quality program amplifier before applying it to the detector. This was done to reduce to a minimum the possible effects of changing signal level on the selectivity of the fork. Tests indicated that no detectable noise or distortion was introduced by the amplifier. The frequency of the test oscillator was monitored and held accurate to better than ±0.2 cycle. The level of the tone applied was adjusted to be 0 db referred to 1 milliwatt at a 0 transmission level point. This is a standard testing level.

To serve as a reference, the results of a noise measurement on a high-grade cable circuit are shown on Fig. 6. This facility gave telewriting performance which was in excellent agreement with the prediction based on *2B* noise meter measurements. This is supported by the curve which closely follows the selectivity curve and approaches the *2B* reading asymptotically. In this case, as well as the others which follow, the reading of the *2B* noise meter was adjusted in accordance with its effective bandwidth to permit comparison with the measuring equipment described in this paper.

The measurements made over a type *O* carrier system are shown on Fig. 7. There appears to be definite evidence of tones superimposed on noise lying in a band extending approximately 15 cycles from the signal. The level of the noise in this band is approximately 20 db higher than the level of noise further removed in frequency from the signal.

A similar picture is obtained from measurements over a type *ON* carrier system as shown on Fig. 8. Here the band of noise extends approximately 30 cycles from the signal.

The results of measurements over a type *N* carrier system are shown on Fig. 9. There is evidence of narrow-band noise extending approximately 30 cycles from the tone. However, its magnitude is not sufficient to explain the telewriter failure. Measurements made over other *O*, *ON*, and *N* systems led to similar results.

Analysis of Results

Detailed examination of these carrier systems leads to several possible explanations of the foregoing results. The type *O* and *ON* carrier systems are single sideband systems employing twin-channel operation,[6] that is, a separate voice channel is derived from each sideband of a channel carrier. The channel carriers are transmitted at depressed levels. At the receiving end the carriers are selected by means of a narrow filter (±40 cycles for type *O* and ±100 cycles for type *ON*), amplified and applied to the demodulator at the proper level to assure stabilized demodulation.

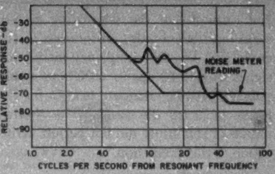


Fig. 8. Noise measurement of type ON carrier telephone system

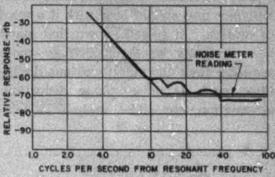


Fig. 9. Noise measurement of type N carrier telephone system

Noise in the band of the pick-off filter consists mainly of carriers of other systems or the carrier shifted slightly in frequency due to inaccuracies of the 304-kc repeater oscillator (in *ON*) which have been superimposed upon the wanted carrier through tertiary crosstalk paths. These unwanted carriers are normally of the order of 40 db below the wanted carrier and so have little effect on the transmission of speech. However, any single frequency transmitted over an *O* or *ON* channel is received with a band of modulation lying on either side of it. Whether this appears as a few discrete frequencies or a band of noise depends on the number of paralleling systems and the number of repeaters (in *ON*). The width of the band of modulation is determined by the differences between the various oscillator crystals involved with an upper limit imposed by the carrier pick-off filter. The foregoing explanation is consistent with the results of the noise measurements over *O* and *ON*. Additional experiments would be required for a definite confirmation.

It may be noted that the level of the narrow-band noise in the type *N* system is approximately 10 db lower than in type *O* and *ON*. This would be expected if the noise were caused by the same mechanism as in *O* and *ON*, since in *N* a higher channel carrier level is employed. However, the magnitude of the noise is insufficient to account for the telewriting performance encountered. Thus, the results of the measurements on *N* carrier indicate that another noise source may be present which was not detected by this method of measurement.

In many cases of analog-data transmission it is possible to overcome the effect of these interferences by increasing the bandwidth used to transmit the signal.

For example, in the telewriting system mentioned previously, an increase in the deviation of the frequency-modulated signal from ±52 cps to ±200 cps would possibly provide adequate margin for the types of noise discussed. The limited nature of this study did not permit verification of this prediction.

Conclusion

The successful use of voice facilities for nonvoice uses should bear with it a word of caution. It is well known that many data systems are considerably more sensitive than the ear to various distortions, such as noise, phase distortion, attenuation, etc. The study of a telewriting system pointed out the need for a technique capable of measuring a narrow band of noise close by in frequency and many db below a signal. Such measurements revealed the presence of a narrow band of interference on either side of a tone transmitted over a few widely used carrier telephone systems. A solution to the problem of transmitting analog signals over these systems is suggested. The design of new analog-data systems for use on voice facilities must take into consideration these small transmission anomalies which are generally of no concern. It is believed that the results of the work just described should prove helpful when considering proposals for sending analog data over voice facilities.

References

1. A. Frequency-Modulation Digital Subset for Data Transmission over Telephone Lines, **Laurance A. Weber.** *AIEE Transactions*, vol. 77, pt. I, 1958 (Jan. 1959 section), pp. 867–72.

2. Recorded Carrier System for High-Speed Data Transmission, **R. M. Gryb.** *Bell Laboratories Record*, New York, N. Y., vol. XXXV, 1957, pp. 321–25.

3. Measurement of Telephone Noise and Power Wave Shape, **J. M. Barstow, P. W. Blye, H. E. Kent.** *Electrical Engineering*, vol. 54, Dec. 1935, pp. 1307–15.

4. Application of Compandors to Message Circuits, **C. W. Carter, Jr., A. C. Dickieson, D. Mitchell.** *AIEE Transactions*, vol. 65, pt. I, 1946, pp. 1079–86.

5. The Type N-1 Carrier Telephone System: Objectives and Transmission Features, **R. S. Caruthers.** *Bell System Telephone Journal*, New York, N. Y., vol. XXX, 1951.

6. Type-O Carrier Telephone, **J. A. Coy, E. K. Van Tassel.** *Electrical Engineering*, vol. 72 May 1953, pp. 418–23.

A Comparison of Mouth-to-Mouth and Manual Artificial Respiration Techniques

ARCHER S. GORDON
NONMEMBER AIEE

CHARLES W. FRYE
NONMEMBER AIEE

CHEMICAL, industrial, and mechanical progress have increased the incidence of cases requiring artificial respiration in industrial, military, and civilian spheres. The danger from electricity is a potent factor in all of these situations.

Recognition of this problem has created a profound interest in resuscitation and has resulted in careful experimental studies of both manual and mechanical methods of artificial respiration. These studies have clearly defined the manual and mechanical techniques which are physiologically superior, and practical application has substantiated these findings.

Throughout this work there has been a growing realization that manual resuscitation and mechanical resuscitation are not distinct entities competing for control of the victim's airway. They are separate but integral parts of the same picture. In the final analysis, there is no conflict between manual artificial respiration and mechanical resuscitation. When available, properly designed and competently employed mechanical resuscitators are easier to use (especially for prolonged periods of time), provide oxygen or air-oxygen mixtures, and leave the operator's hands free for maintaining an unobstructed airway. Such devices should be brought into operation as soon as practical. However, the use of mechanical resuscitators must remain always a secondary consideration. In resuscitation emergencies, time is of the essence, seconds count, and the breach between respiratory failure and the use of mechanical devices must always be filled by the use of a manual technique (Fig. 1). However, it is not sufficient merely to satisfy this time requirement by immediately beginning manual artificial respiration. It is equally important: 1. to use the best manual method; 2. to perform the method correctly and adequately; and 3. to maintain an unobstructed airway for unimpeded ventilation.

Background for Current Studies

Several years ago a series of carefully controlled human studies was carried out to evaluate the various factors concerned with manual artificial respiration. These tests were performed on warm, nonrigid corpses,[1] anesthetized normal adult male volunteers whose breathing was stopped with drugs,[2] surgical patients,[3] and patients whose breathing stopped as a result of various disease processes.[4] Ventilatory data indicated that "push-pull" manual methods, which have an active inspiratory phase and an active expiratory phase, are the most effective. These include: back-pressure arm-lift (modified Holger Nielsen); back-pressure hip-lift (Schafer-Emerson-Ivy); and arm-lift chest-pressure (Silvester), as illustrated in Fig. 2.

Circulatory studies[5] also were performed and revealed that all of these methods maintained near-normal amounts of oxygen in the blood. Air-flow studies[6] indicated that a performance rate of approximately 12 complete cycles per minute provided optimum ventilation. Performance and fatigue tests[7–10] showed the back-pressure hip-lift method to be more tiring and less accurately performed than back-pressure arm-lift. Airway obstruction, it was noted, occurred in 50% of the cases using the Silvester method due to the supine position.[1]

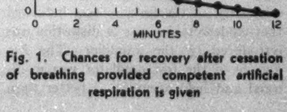


Fig. 1. Chances for recovery after cessation of breathing provided competent artificial respiration is given

---

Paper 57-1012, recommended by the AIEE Safety Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted June 17, 1957; made available for printing December 17, 1958.

Archer S. Gordon, M.D., and Charles W. Frye, M.D., are with the University of Illinois College of Medicine, Chicago, Ill.

Vol.
Pt. 1
1959

On the basis of these studies, the National Research Council recommended in 1951 that the back-pressure arm-lift method was the preferred technique for general use. The other two push-pull methods were recommended as useful where particular circumstances demanded their application.[11] This decision was accepted by the American Red Cross, American Medical Association, Department of Defense, U.S. Public Health Service, Federal Civil Defense Administration, and most other national and federal agencies, utilities, and organizations concerned with resuscitation.

Pole-Top Method

It was suggested subsequently that the push-pull principle be applied to pole-top resuscitation. Human studies were performed[12] and revealed the ventilatory superiority of this method (Fig. 3) over the standard pole-top technique.[13,14] On a practical basis, push-pull pole-top resuscitation has been found effective. However, it has been criticized in some instances because the situation of the victim in relation to crossbars, wires, etc., precluded proper and adequate performance of the maneuvers.

A worthwhile recommendation for such situations (suggested in a personal communication from T. M. C. Martin) would be to have mechanical resuscitators equipped with extension hoses from the tanked gases to the face mask and all possible operating controls at, or near, the mask. This would allow rapid, adequate, and easy application on the pole. The mechanical resuscitation could be continued while the victim was being lowered. This practical adaptation of mechanical resuscitation deserves further investigation and development. Useful as such equipment might be, however, the urgency still remains for immediate ventilation of the victim. Accordingly, until such a mechanical device can be taken up the pole to the victim, some form of manual artificial respiration should be applied. Where the situation allows use of push-pull maneuvers these are preferable. If not feasible, standard pole-top resuscitation should be used.

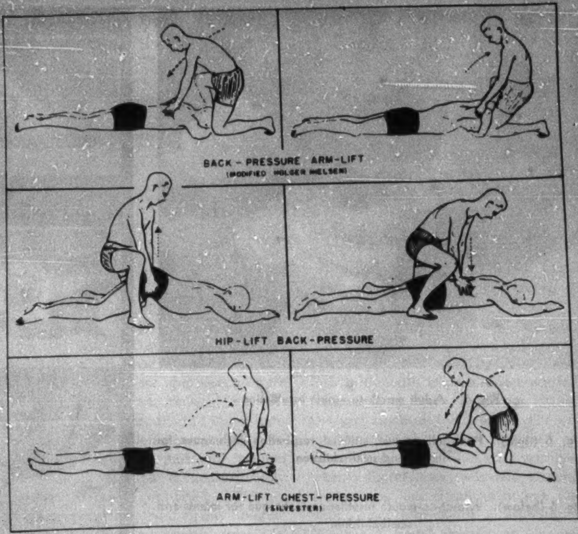

BACK - PRESSURE ARM-LIFT
(MODIFIED HOLGER NIELSEN)
HIP - LIFT BACK - PRESSURE
ARM - LIFT CHEST - PRESSURE
(SILVESTER)

Fig. 2 (above). Push-pull methods of manual artificial respiration

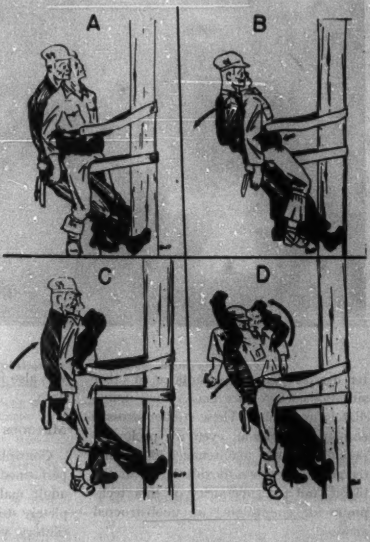


Fig. 3 (right). Push-pull pole-top resuscitation

Evaluation of Mouth-to-Mouth Method

In October 1955, Rickard[15] suggested that infants and small children could best be resuscitated manually by rocking them alternately into a 45-degree head-down and 45-degree feet-down position on the forearm of the rescuer. This modified Eve method[16] included the use of a digital airway at all times provided by the operator's middle finger which was inserted into the child's mouth to keep the jaw and tongue from obstructing the airway. The importance of maintaining a clear airway was stressed.

Concurrently, in the Armed Forces, Johns and Cooper,[17–18] and Elam and Brown, and associates,[19–21] were studying mask-to-mask insufflation for personnel requiring resuscitation in an atmosphere contaminated with nerve gases. They also studied mouth-to-mask insufflation

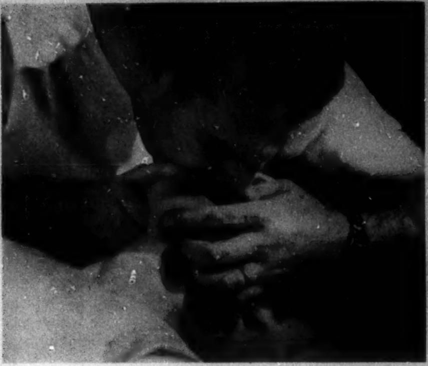
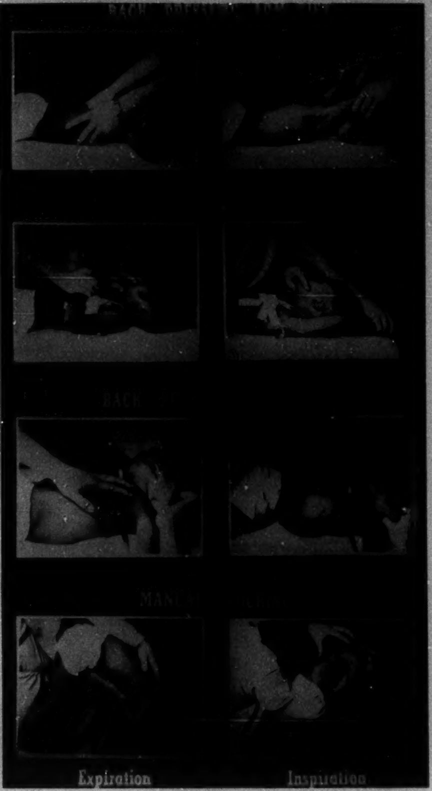


Fig. 4. Adult mouth-to-mouth insufflation

Fig. 5 (right). Push-pull manual artificial respiration techniques for infants and small children

Fig. 6 (below). Mouth-to-mouth insufflation technique for infants and small children



for use on hospital patients or in other situations where face masks were immediately available. These studies stressed ease of application, even for prolonged periods of time, maintenance of normal oxygen and carbon-dioxide levels in the blood, and the superiority of this technique for maintaining an unobstructed airway.

Since all of these recent studies have stressed the factor of airway patency, the authors undertook to evaluate this factor with the manual push-pull methods and direct mouth-to-mouth insufflation on a series of nonbreathing adults and on a similar series of infants and small children. The manual-rocking method was also included in the survey on children.

Methods

Controlled, comparative studies were performed on eight anesthetized normal adult males whose breathing was completely stopped by means of drugs. Ventilatory volumes were measured during the application of mouth-to-mouth insufflation, Silvester, back-pressure arm-lift, and back-pressure hip-lift methods of resuscitation. These were performed first with a cuffed endotracheal tube in the airway. This indicated the volumes achieved under ideal circumstances with the patency of the airway assured. The methods were then repeated without the endotracheal tube. Instead, a rubber mouthpiece was fitted between the teeth and lips and a clip was used to close off the nose. The mouthpiece was then connected to the spirometer to measure ventilatory volume. This leaves the tongue and jaw free to assume their normal position during actual resuscitation maneuvers and provides a measure of the airway obstruction occurring with each of the methods. The manual methods were performed as illustrated in Fig. 1, and mouth-to-mouth insufflation as shown in Fig. 4.

The infant studies were carried out on patients ranging in age from 5 to 36 months. They were performed prior to elective surgical procedures in a manner similar to the adult tests. The manual methods were applied both with and without a digital airway wherever pos-

Table I. Mean Pulmonary Ventilation During Artificial Respiration on Eight Apneic Normal Adult Male Subjects

Methods	Cuffed Endotracheal Tube: Cubic Centimeter per Respiratory Cycle	Mouthpiece and Noseclip: Cubic Centimeter per Respiratory Cycle	Mouthpiece and Noseclip: % of Cases Actually Ventilated*
Normal resting tidal volume	640		
Mouth-to-mouth	980	890	100%
Silvester	820	450	50%
Back-pressure arm-lift	950	580	75%
Back-pressure hip-lift	1090	650	75%

* In similar previous studies,[1] using mouthpiece and noseclip on eight warm, nonrigid corpses, the per cent of cases actually ventilated was: Silvester, 50%; back-pressure arm-lift, 100%; back-pressure hip-lift, 100%.

sible. Since the back-pressure arm-lift method requires the use of two hands, it was impossible for a single operator to insert a digital airway with this technique. The push-pull manual methods and manual rocking are illustrated in Fig. 5, and mouth-to-mouth insufflation for infants in Fig. 6.

RESULTS

The ventilatory data from the adult studies are summarized in Table I. With an ideal airway provided by the cuffed endotracheal tube, all push-pull manual methods and mouth-to-mouth insufflation gave the same range of ventilation, and this was one quarter to three quarters more than the resting tidal volume. When the mouthpiece and noseclip were used, ventilation was reduced with all methods. However, mouth-to-mouth insufflation showed less than a 10% decrease, whereas the push-pull manual methods all resulted in approximately 40% to 50% less ventilation. This reduction is due to partial airway obstruction without the endotracheal tube. Of even greater importance is the observation that all cases were ventilated by mouth-to-mouth insufflation, whereas only 75% of them were ventilated by the back-pressure arm-lift and back-pressure hip-lift techniques, and only 50% with use of the supine Silvester method.

The data on infants and small children are itemized in Tables II and III. With the endotracheal tube providing an ideal airway, mouth-to-mouth insufflation always provided maximum pulmonary ventilation, usually two to three times the normal resting tidal volume. The manual push-pull techniques all provided the same range of ventilation and usually about one half to two thirds as much as the mouth-to-mouth method. All variations of manual rocking were significantly inferior and resulted in less ventilation than the resting tidal volume. When the mouthpiece and noseclip were used, mouth-to-mouth resuscitation essentially resulted in as much ventilation as with the endotracheal tube, and every case was ventilated. With the push-pull manual methods all cases were ventilated when a digital airway was used, although the volume was significantly less than with mouth-to-mouth resuscitation. When the digital airway was not used, only one half of the cases were ventilated. With the manual rocking techniques, only one half of the patients were ventilated

Table II. Pulmonary Ventilation During Artificial Respiration on Five Apneic Infants

Infant: Number	Infant: Age, Months	Infant: Weight, Pounds	Sleeping Tidal Volume	Tidal Volume With Endotracheal Intubation (Cubic Centimeter Per Respiratory Cycle): Mouth-to-Mouth	Silvester	Back-Pressure Arm-Lift*	Back-Pressure Hip-Lift†	Manual Rocking: Prone: 45 to 45 degrees	Manual Rocking: Prone: 30 to 60 degrees	Manual Rocking: Supine: 45 to 45 degrees	Manual Rocking: Supine: 30 to 60 degrees
1	5	15	32	94	46	41	58	18	20	22	25
2	12	22	48	150	98	62	106	24	28	30	36
3	24	31	62	220	110	86	116	35	42	45	50
4	30	30	87	248	124	111	133	42	46	52	62
5	36	38	94	284	206	184	210	65	72	86	92

* In all cases, arm-lift alone gave values of 30 to 50% of back-pressure arm-lift.
† In all cases, hip-lift alone gave values of 35 to 45% of back-pressure hip-lift.

Table III. Pulmonary Ventilation During Artificial Respiration on Nine Apneic Infants

Infant: Number	Infant: Age, Months	Infant: Weight, Pounds	Sleeping Tidal Volume	Tidal Volume With Mouthpiece and Noseclip† (Cubic Centimeter Per Respiratory Cycle): Mouth-to-Mouth	Silvester: (A) With	Silvester: (B) Without	Back-Pressure Arm-Lift: (B) Without	Back-Pressure Hip-Lift: (A) With	Back-Pressure Hip-Lift: (B) Without	Manual Rocking: Prone (45 to 45 degrees) (A) With	Manual Rocking: Supine (45 to 45 degrees) (A) With
1*	5	15	32	86	16	0	0	26	0	0	0
2*	12	22	48	140	58	40	81	131	106	10	6
3	17	25	56	122	13	0	0	38	0	0	0
4	18	23	58	148	58	38	44	88	72	20	0
5*	24	27	62	134	40	30	22	81	40	0	0
6	28	32	70	184	66	0	0	87	0	0	0
7*	30	31	87	196	25	0	0	140	88	26	0
8	32	36	88	220	16	0	0	47	0	13	20
9*	36	38	94	260	62	78	16	87	30	0	0

* Same cases as in Table 1.
† In each series, "(A) With" indicates the results obtained when the method was performed with the operator's finger in the mouth to keep the airway clear; "(B) Without" indicates the results without a finger in the airway; back-pressure arm-lift requires the operator to use both hands, therefore, all of these measurements were made without a finger in the airway; all manual rocking was performed with a finger in the airway.

even with a digital airway. This was attributed to flexion of the neck during rocking which frequently resulted in obstruction of the airway.

DISCUSSION

Direct mouth-to-mouth resuscitation has been recommended and practiced since Biblical days. But comparative, controlled tests have never been performed previously. They emphasize a number of important features of the technique.

With an assured ideal airway, mouth-to-mouth insufflation (at an optimum pressure of approximately 15 millimeters mercury) provides the same tidal ventilation as a well-performed push-pull manual method. However, when an endotracheal tube is not used, airway obstruction becomes a vital problem which can partially or completely nullify the resuscitation effects. Properly performed mouth-to-mouth insufflation is the only nonmechanical technique which provides optimal ventilation in all cases for both infants and adults. Carefully performed push-pull manual methods provide only partial or sometimes no ventilation, unless a second rescuer can devote himself to assuring the patency of the airway, or, for children, a digital airway is used.

There are other advantages to mouth-to-mouth insufflation. When the rescuer breathes into the victim's airway, he can immediately detect obstruction by feeling resistance to insufflation and by noting that the chest fails to rise and fall rhythmically with each breath. Since he is stationed at the victim's head, he easily can check the airway to relieve the obstruction. With none of the manual methods is the operator able to detect airway obstruction nor is he in a good position to relieve it.

If partial obstruction occurs and cannot be digitally relieved, the rescuer can increase his insufflation pressure in order to ventilate the lungs. This is not possible with the manual methods. He can also slow his rate of insufflation to bypass a partial obstruction. Additional pressure is also useful for ventilating the noncompliant lung, such as is found in cases associated with pulmonary edema. Such pressure increases cannot be accomplished with any manual method.

Elam's studies reveal the ease and effectiveness of mouth-to-mouth insufflation even for long periods of time. In addition, disproportions in size do not limit the small operator from performing mouth-to-mouth insufflation on a large victim as do the push-pull manual methods.

Inspired air contains 20 to 21% oxygen and a trace of carbon dioxide; expired air is approximately 16% oxygen and 4 to 5% carbon dioxide. This amount of oxygen is more than adequate for the body's needs, and this low concentration of carbon dioxide is not deleterious if sufficient pulmonary ventilation is achieved. Elam's studies[19-21] have shown conclusively that mouth-to-mouth insufflation can maintain normal amounts of oxygen and carbon dioxide in the victim's blood during long periods of application.

Mouth-to-mouth insufflation can be applied in many situations where the manual methods are seriously limited, such as for fractures or severe injuries of extremities and thorax, for burns, and for cases where the victims are in unusual positions or parts of their bodies are trapped. Whether it would be practical to attempt mouth-to-mouth ventilation of a pole-top victim is not known. There are times when it might be possible and preferable. Rather than advocate one or the other method, it would appear wiser to re-emphasize that some method must be started at once. The individual operators should be capable of performing the mouth-to-mouth technique as well as standard pole-top and push-pull pole-top maneuvers. Mouth-to-mouth and push-pull pole-top resuscitation give greater ventilation and should be used when possible. Otherwise standard pole-top resuscitation should be started and continued until the victim can be lowered to the ground or a mechanical resuscitator with an extension mask brought up to him on the pole.

RECOMMENDATIONS AND TECHNIQUE

As a result of these recent findings, the National Research Council and the American Red Cross have recommended that the preferable nonmechanical method of resuscitation for infants and small children is mouth-to-mouth insufflation. They also state that this method may be used for adults. Although the efficacy of the method has been recognized for adults, it is felt by some that the aesthetic aspects of performing mouth-to-mouth insufflation on near-dead, cyanotic, moist, etc., adult victims may deter rescuers from using it. In such cases, push-pull manual methods remain as an alternative. To overcome some of these aesthetic considerations, a clean piece of cloth or gauze may be interposed between the rescuer's and victim's mouths. However, this makes performance more difficult, somewhat increases the resistance to ventilation, and may not be readily available. Valuable time should not be lost in seeking or applying this adjunct.

The technique for performing mouth-to-mouth insufflation is essentially as illustrated in Figs. 4 and 6. For infants and small children, the rescuer should place the victim on his back and use the middle and/or ring fingers of both hands to lift the lower jaw from beneath and behind so that it juts out. This draws the base of the tongue forward and keeps the airway from obstructing. He then places his mouth over the child's mouth *and* nose, making a relatively leakproof seal, and breathes into the child's airway with a smooth, steady action until he observes the chest rise. This phase is continued until the rescuer feels the resistance of the victim's lungs. As he starts this action, he moves one of his hands to the child's stomach, between the navel and the ribs, and applies continuous, moderate pressure to prevent the stomach from becoming distended with air. The other hand is maintained in its position under the jaw to keep the airway clear. As soon as the lungs have been inflated, the rescuer removes his lips from the child's mouth and nose and allows the lungs to empty passively. The cycle is then repeated, keeping one hand behind and beneath the jaw, and the other pressing on the stomach at all times. This should be continued at a rate of about 20 per minute. After every 20 cycles, the operator should rest long enough to take one deep breath. If at any time he feels resistance to his breathing into the airway, or fails to see the chest rise and fall rhythmically, he should quickly lift the lower jaw again with both hands in order to keep the airway clear.

If the rescuer cannot ventilate the lungs of the victim, even after carefully holding the jaw forward as directed, the possibility of a foreign body in the airway should be entertained. The victim should be placed on the rescuer's forearm in a face-down head-down position with the middle finger of the rescuer's hand inserted into the airway to hold the tongue forward. He should then pat him firmly on the back to dislodge any foreign body. Mouth-to-mouth insufflation should then proceed as directed.

The method is essentially the same for adults, with the following exceptions. The rescuer breathes only into the mouth (or nose) of the victim. Both hands are usually required to keep the jaw elevated and it is not always possible to move one hand to the abdomen. The rate should be approximately 12 complete cycles per

Vol.
Pt. 1
195

minute, with the rescuer taking one deep breath between each 12 cycles. When the rescuer breathes into the victim's mouth, it is necessary for him to obstruct the nose with his two thumbs in order to prevent the escape of air.

Summary

1. With an ideal, unobstructed airway, mouth-to-mouth insufflation has been found to give as effective ventilation as push-pull manual methods of artificial respiration.

2. When an ideal airway is not assured, such as under practical resuscitation conditions, the mouth-to-mouth method is the only nonmechanical technique which assures a patent airway.

3. Even the well-performed push-pull manual methods frequently result in partial or complete obstruction of the airway.

4. Mouth-to-mouth insufflation is the preferred nonmechanical method of resuscitation for infants and small children. It should also be used for adults if the rescuer can overlook the aesthetic aspects of the situation. Push-pull manual methods are a second choice.

5. For pole-top resuscitation, some method of artificial respiration must be started at once. Mouth-to-mouth insufflation and push-pull pole-top resuscitation give the best ventilation and should be used if practical. If not, standard pole-top resuscitation should be employed until a mechanical resuscitator with mask extension can be taken up the pole, or the victim can be lowered to the ground.

References

1. Artificial Respiration: A New Method and a Comparative Study of Different Methods in Adults, **A. S. Gordon, D. C. Fainer, A. C. Ivy.** *Journal*, American Medical Association, Chicago, Ill., vol. 144, 1950, p. 1455.

2. Manual Artificial Respiration: Comparison of Effectiveness of Various Methods on Apneic Normal Adults, **A. S. Gordon, F. Raymon, M. Sadove, A. C. Ivy.** *Ibid.*, p. 1447.

3. Comparison of Methods for Performing Manual Artificial Respiration, **R. G. Nims, E. H. Conner, S. Y. Botelho, J. H. Comroe, Jr.** *Journal of Applied Physiology*, Washington, D. C., vol. 4, no. 6, Dec. 1951, p. 486.

4. Mechanics of Breathing in Relation to Manual Methods of Artificial Respiration, **J. L. Whittenberger, J. E. Affeldt, W. T. Goodale, S. J. Sarnoff.** *Ibid.*, p. 476.

5. Circulatory Studies During Artificial Respiration on Apneic Normal Adults, **A. S. Gordon, O. Prec, H. Wedell, M. S. Sadove, R. Raymon, J. T. Nelson, A. C. Ivy.** *Ibid.*, p. 421.

6. Air-Flow Patterns and Pulmonary Ventilation During Manual Artificial Respiration on Apneic Normal Adults, **A. S. Gordon, J. E. Affeldt, M. S. Sadove, F. Raymon, J. L. Whittenberger, A. C. Ivy.** *Ibid.*, p. 408.

7. Energy Expended in Administering Artificial Respiration, **P. V. Karpovich, C. J. Hale.** *Ibid.*, p. 467.

8. Manual Artificial Respiration: Pedagogical and Fatigue Factors Involved in Its Use, **P. V. Karpovich, C. J. Hale.** *Ibid.*, p. 472.

9. Energy Expenditure of Operators During Manual Artificial Respiration, **A. S. Gordon, F. Raymon, M. S. Sadove, H. Wedell, A. C. Ivy.** *Ibid.*, p. 439.

10. Pedagogical and Performance Factors of Manual Artificial Respiration with Naval Personnel, **A. S. Gordon, S. Star, F. Meier, C. J. Hale, A. C. Ivy.** *Ibid.*, p. 447.

11. Critical Survey of Manual Artificial Respiration, **A. S. Gordon, M. S. Sadove, F. Raymon, A. C. Ivy.** *Journal*, American Medical Association, vol. 147, Dec. 1951, p. 1444.

12. New Push-Pull Pole-Top Resuscitation, **A. S. Gordon, C. W. Frye, M. S. Sadove.** *Industrial Medicine and Surgery*, Chicago, Ill. vol. 21, no. 7, July 1952, pp. 326–32.

13. Using Vital Movements in Saving a Life, **E. A. Oesterreich.** *NELA Bulletin 228*, National Electric Light Association, Apr. 1932.

14. Pole-Top and Other Manual Resuscitation Methods, **A. S. Gordon, M. S. Sadove, F. Raymon, A. C. Ivy.** *Industrial Medicine and Surgery*, vol. 21, no. 4, Apr. 1952, pp. 147–52.

15. A New Method of Manual Artificial Respiration for Infants and Small Children, **H. J. Rickard.** *Journal*, American Medical Association, vol. 159, no. 8, 1955, p. 754.

16. Activation of the Inert Diaphragm by a Gravity Method, **F. C. Eve.** *The Lancet*, London, England, vol. 2, Nov. 5, 1932, p. 995.

17. **R. J. Johns, D. Y. Cooper.** *Medical Research Report no. 76*, Chemical Corps, Army Chemical Center, Edgewood, Md., Aug., 1951.

18. A Mask-to-Mask Resuscitator Proposed for Chemical Corps Use, **D. Y. Cooper.** *MLSR no. 58*, Army Chemical Center Medical Laboratory, Dec. 1954.

19. Resuscitation With Mouth-to-Mouth Methods. Artificial Ventilation for Emergency and Prolonged Periods in Normal or Contaminated Atmospheres, **E. S. Brown, J. O. Elam.** *Report no. 28*, Army Chemical Center Medical Laboratory, July 1953.

20. Artificial Respiration by Mouth-to-Mask Method, **J. O. Elam, E. S. Brown, J. D. Elder, Jr.** *New England Journal of Medicine*, Boston, Mass., vol. 250, no. 18, 1954, p. 749.

21. Mask-to-Mask Studies in Paralyzed Human Subjects, **J. O. Elam, E. S. Brown, J. A. Clements, D. G. Greene, R. O. Bauer, C. D. Janney.** *CWLR 2082*, Army Chemical Center Medical Laboratory, Dec. 13, 1956.

Thermal Life of Enameled Magnet Wire

AIEE COMMITTEE REPORT

IN THE SUMMER of 1953, a Working Group on Thermal Stability of Insulating Materials was organized under the sponsorship of the AIEE Subcommittee on Electrical Insulation (formerly the Dielectrics Committee) of the Basic Sciences Committee. In January 1954 a task force on the Evaluation of the Thermal Stability of Enameled Magnet Wire was appointed and it was given the responsibility of developing a test procedure.

A suggested test method presented at the 1955 AIEE Winter General Meeting was adopted for further study and a cooperative test program was started.[1] This method consisted of heat aging National Electrical Manufacturers Association electric strength test samples and subjecting them to a predetermined test voltage applied periodically during the aging. The end point was determined by the breakdown of the wire specimens.

As a result of this program, a report and a proposed test procedure were presented at the AIEE Fall General Meeting in October 1954. In October 1955, AIEE Publication no. *57* was issued by the Institute.[2] This test procedure is the responsibility of the Working Group on Enameled Magnet Wire, Subcommittee on Thermal Evaluation, Electrical Insulation Committee.

Temperature Classification

The 1957 revision of AIEE Publication no. *1* establishes the concept that temperature ratings of insulating materials and systems shall be based upon "experience or accepted tests."

Publication no. *1* further states that the classification of insulating materials involves two steps: 1. preparing test procedures, and 2. establishing minimum thermal life values based upon results of aging tests on known materials.

Although AIEE no. *1* specifies that

Paper 58-1205, recommended by the AIEE Electrical Insulation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958, and represented at the Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted June 27, 1958; made available for printing August 18, 1958.

The members of the Working Group on Enameled Magnet Wire, Subcommittee on Thermal Evaluation of the AIEE Electrical Insulation Committee are J. F. Dexter, *chairman*, Ralph Hall, K. N. Mathes, E. L. Brancato, H. L. Saums, F. A. Sattler, D. L. McClenahan, J. M. Keller, R. B. Young, T. J. Gair, F. E. Dresner, W. W. Wareham, C. E. Logston, Edward Lavin, K. Alexander.

technical committees will determine the minimum thermal life values for new materials, temperature classification involves commercial considerations in addition to purely technical questions. In trying to develop a procedure for classifying the temperature ratings of enameled magnet wire, it became obvious that these commercial matters could not be resolved by the AIEE Working Group. However, it is important that some practical procedure be worked out so that the temperature rating of each new insulating material can be determined.

As a step toward the preparation of these specifications, thermal-aging data must be available. For the past several years, the Electrical Insulation Committee has sponsored the presentation of technical papers on magnet-wire thermal-stability studies. During this time, members of the AIEE Working Group have obtained data on the many different types of enameled wire. These data may be used to establish base lines against which the life of new enamels be compared.

Test Procedure

Data given in this report have been obtained according to the methods outlined in AIEE no. *57* test procedure. Deviations from no. *57* are noted where they are known to exist.

Data are given on unvarnished and varnished test specimens. The general chemical classification of the insulating varnishes used to coat the specimens is given when available, but formulations within a class can vary widely. In many cases, the wire size was not reported. However, most of these data were obtained on no. *18* Awg (American Wire Gauge) and no. *20* Awg copper wire.

Types of Enameled Magnet Wire

In reporting thermal life data, commercial enamels are classified according to their chemical composition. Generally, each type of enamel has a thermal life-temperature characteristic typical of its chemical composition, but variations in life can be expected for different formulations within a chemical type.

The following chemical classifications have been selected arbitrarily in presenting data:

Polyvinyl formal
Epoxy
Epoxide-polyester
Polyester
Polyester-polyamide
Polyamide

Table I. Statistical Analyses of Thermal Life Data

		95% Confidence Limits, Hours			
Temperature, C	Life, Hours	Upper Regression	Lower Regression	Upper Data	Lower Data
Polyvinyl Formal, Unvarnished Specimens Coefficient of Multiple Correlation (CMC): $r^2=92\%$					
105	15,011	20,571	10,954	30,848	7,305
130	3,341	4,086	2,732	6,583	1,695
155	886	994	790	1,711	459
180	272	298	248	523	141
220	53	63	44	103	27
Polyvinyl Formal, Varnished Specimens CMC: $r^2=87\%$					
105	31,592	54,424	18,339	77,724	12,841
130	6,697	9,540	4,701	14,903	3,009
155	1,701	2,076	1,394	3,582	808
180	502	567	445	1,041	242
220	92	118	92	197	43
Epoxy, Unvarnished Specimens CMC: $r^2=85\%$					
105	46,242	108,782	19,657	160,060	13,359
130	9,785	17,540	5,459	28,603	3,347
155	2,482	3,549	1,736	6,538	942
180	732	901	595	1,845	290
220	134	181	99	347	52
Epoxy, Varnished Specimens CMC: $r^2=72\%$					
105	53,697	115,249	25,019	204,122	14,126
130	13,559	23,272	7,900	45,990	3,997
155	4,021	5,702	2,835	12,696	1,273
180	1,363	1,666	1,116	4,152	447
220	303	365	252	922	99
Epoxide-Polyester, Unvarnished Specimens CMC: $r^2=71\%$					
105	168,568	487,624	58,272		28,319
130	28,118	60,888	12,985	143,236	5,520
155	5,781	9,918	3,370	26,735	1,250
180	1,415	2,080	962	6,242	321
220	200	296	135	885	45
Epoxide-Polyester, Varnished Specimens CMC: $r^2=84\%$					
105	212,861	392,917	115,317	657,157	68,948
130	36,520	56,062	23,790	103,180	12,926
155	7,698	10,145	5,841	20,624	2,873
180	1,926	2,288	1,622	5,040	736
220	281	345	229	740	106
Polyester, Unvarnished Specimens CMC: $r^2=85\%$					
105					
130	300,726	578,096	156,438	889,580	101,662
155	41,411	64,859	26,440	109,777	15,621
180	7,097	9,314	5,407	17,582	2,864
220	612	682	549	1,464	256
Polyester, Varnished Specimens CMC: $r^2=86\%$					
105					
130	276,820	429,069	178,594	812,890	94,267
155	44,375	60,396	32,604	124,447	15,823
180	8,706	10,606	7,146	23,751	3,191
220	906	993	826	2,434	337
Polyamide, Unvarnished Specimens CMC: $r^2=87\%$					
105	22,306	52,051	9,559	69,055	7,205
130	5,314	9,321	3,030	13,540	2,085
555	1,497	2,077	1,078	3,386	661
180	485	589	399	1,050	224
220	101	144	71	231	44
Acrylic, Unvarnished Specimens CMC: $r^2=87\%$					
105	44,182	80,103	24,369	130,749	14,930
130	7,399	10,982	4,985	19,903	2,751
155	1,526	1,925	1,210	3,895	598
180	375	432	325	939	140
220	53	68	41	136	20
Acrylic, Varnished Specimens CMC: $r^2=60\%$					
105	75,852	299,686	19,198	600,896	9,574
130	12,971	34,234	4,914	80,610	2,087
155	2,726	5,080	1,463	14,457	514
180	680	956	484	3,320	139
220	99	132	73	478	20
Polyester-Polyamide, Varnished Specimens CMC: $r^2=86\%$					
105	28,729	48,378	17,061	70,292	11,742
130	9,635	14,190	6,542	21,963	4,227
155	3,671	4,839	2,785	7,992	1,686
180	1,516	1,888	1,281	3,303	733
220	472	554	402	994	224

Table I. (Cont).

Temperature, C	Life, Hours	95% Confidence Limits, Hours			
		Upper Regression	Lower Regression	Upper Data	Lower Data
Polyurethane, Unvarnished Specimens* CMC: $r^2=0.63\%$					
Polyurethane, Varnished Specimens CMC: $r^2=43\%$					
105	96,173		8,676		3,869
130	15,374	79,681	2,966	226,848	1,041
155	3,044	8,221	1,127	31,936	290
180	720	1,181	439	6,420	80
220	97	189	50	907	10
Polyester Modified Silicone, Unvarnished Specimens CMC: $r^2=88\%$					
105					
130					
155	413,642	830,145	206,108		113,720
180	43,171	70,009	26,622	141,892	13,135
220	1,870	2,312	1,512	5,662	617
250	243	282	209	729	81
Polyester Modified Silicone, Varnished Specimens CMC: $r^2=94\%$					
105					
130					
155	352,887	522,489	238,338	760,541	163,737
180	39,867	52,367	30,350	81,426	19,519
220	1,928	2,169	1,713	3,769	986
250	269	290	249	523	138

* Computations indicate that the thermal life data do not define a single regression line.

Acrylic
Polyurethane
Silicone modified polyester
Polytetrafluoroethylene

Statistical Analysis of Thermal Life Data

Thermal life values for each type of enamel are presented in tabular form. Each value given in the tables is the average life of ten or more magnet wire specimens. Life values for varnished and unvarnished specimens are given. The equation for the regression line for each enamel was determined according to the procedure given in AIEE no. *1(F)*.

In Table I, a detailed summary of the statistical calculations is given. These data may be useful to those wishing to reproduce the graphs or to compare thermal life values of different enamels.

In analyzing the data obtained on varnished specimens, the variations in life caused by different varnishes must be considered. Since different varnishes may result in different life–temperature relationships, the regression line is of little significance. Therefore, with two exceptions, graphs showing the life as a function of aging temperature are given only for unvarnished specimens.

As a result of the AIEE test programs, there were considerable data available for polyvinyl formal coated with a phenolic-alkyd varnish and for a silicone modified polyester enamel coated with a silicone varnish. In these cases a graph is included showing the life temperature characteristic of these varnished specimens.

Thermal Life of Enameled Magnet Wire

Polyvinyl Formal Enameled Wire. In Table II, thermal life values are given for unvarnished specimens of copper wire enameled with commercial polyvinyl formal enamels. Thermal life values obtained on varnished specimens are given in Table III. Most of the data on varnished specimens were obtained with modified alkyd varnish. However, the type of varnish used may affect significantly the thermal life. For example, in Table III the thermal life values reported by laboratory *E* show the difference between specimens dipped in polyester varnish and in phenolic varnish.

The equation for the regression line and the 95% confidence limits on the estimated life at 105 C (degrees centigrade) for unvarnished specimens are given in Table II. The range of 10,954 to 20,571 hours for 95% confidence in the averge life at 105 C is narrow, considering that these data were obtained over several years on different commercial enamels by nine laboratories. It is doubtful that further testing of this type of enamel will significantly improve the estimated life at 105 C. Data given in Tables II and III are shown graphically in Figs. 1 and 2.

This enamel has been used successfully for many years in electric equipment designed for class *A* temperature rise. Therefore, it is logical to select polyvinyl formal enamel as the "accepted material," and its thermal life taken as a base for comparing the relative life of new enamels.

Epoxy Enameled Wire. Thermal life values for unvarnished and varnished specimens of epoxy enameled wire are given in Tables IV and V. In Fig. 3, the thermal life–temperature characteristic of unvarnished specimens is shown.

Epoxy-Polyester Enameled Wire. The thermal life values for unvarnished and varnished specimens of a polyester modified epoxide wire enamel are given in Tables VI and VII. The life as a function of temperature for this enamel is shown in Fig. 4.

Polyester Enameled Wire. Thermal life values for unvarnished and varnished

(*Text continues on page 101*)

Table II. Thermal Life Polyvinyl Formal Enameled Magnet Wire, Unvarnished Specimens

Laboratory	Life, Hours												
	125 C	130 C	140 C	150 C	160 C	170 C	175 C	180 C	190 C	200 C	210 C	220 C	225 C
A		2,670		735		256	347		72		66		
				925		375	323		135		67		
				910		178			101				
				1,545									
B	6,800			1,240	810		402			115		43	
				1,350	828					128		44	
										122			
C				910			460			109			
D		3,800			680			378		228			
		3,470			640			265		155			
					720			365		202			
					588			410		155			
E			2,484	815	698			197		119			
					552			303		91			
					640			341		92			
								309		96			
F					1,285			347		135			
					790			333		112			
G								300*		150*			75*

Equation of regression line: Log life $= -6.3457 + 3{,}979/C + 273$.

Estimated life at 105 C: 15,011 hours.

95% confidence limits at 105 C: 10,954 to 20,571.

* Failure criterion 2 kv; data not included in calculating average life.

Table III. Thermal Life Polyvinyl Formal Enameled Magnet Wire, Varnished Specimens

Laboratory	Varnish Type	Life, Hours 130 C	140 C	150 C	160 C	170 C	175 C	180 C	190 C	200 C	210 C	220 C	225 C
A	Phenolic	4,070		1,974		482	681		267		104		
	Alkyd			955		739			419		145		
				2,790		1,067			316		120		
						1,090			315		127		
						617			294		151		
						1,560							
B	Phenolic						1,130			210			67
	Alkyd									192			
C	Phenolic						1,200			207			
	Alkyd												
D	Epoxy†											231	
D	Phenolic				967			358					
	Alkyd												
D	Polyester†				2,551			1,323		1,291		406	
										764			
D	Silicone†							1,194					
	Alkyd												
E	Polyester†		5,832		3,871			2,366		741		492	
E	Phenolic†		3,348		1,097			471		213			
F	Phenolic				3,216			763		143			
	Alkyd				1,411			557		171			
G	Phenolic			4,500*		1,000*				280*			90*
	Alkyd												
I	Phenolic									165			54

Equation of regression line: Log life $= -6.3643+4,108/C+273$.

* Failure criterion 2 kv; data not used in calculating average life.

† Not included in statistical study.

Table IV. Thermal Life Epoxy Enameled Magnet Wire, Unvarnished Specimens

Laboratory	Life, Hours 150 C	170 C	175 C	180 C	190 C	200 C	210 C	220 C	225 C	230 C	250 C
A	2,800	1,400	660		552	215	355			65	
	3,052		631			106					
B			1,210			418		319	172		
C	3,720		970			215			49		
D				1,172		415					
						307					
						308					

Equation of regression line: Log life $= -6.2112+4,113/C+273$.

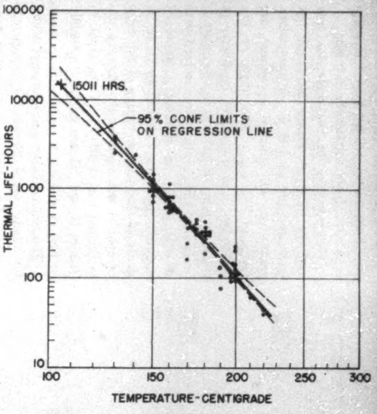


Fig. 1. Thermal life of polyvinyl formal enameled magnet wire, unvarnished specimens. Data in figures from AIEE no. 57 test procedure

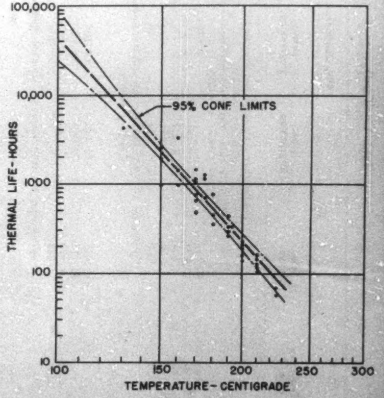


Fig. 2. Thermal life of polyvinyl formal enameled magnet wire, varnished specimens

Vol.
Pt.
195

Table V. Thermal Life Epoxy Enameled Magnet Wire, Varnished Specimens

Laboratory	Varnish Type	Life, Hours 150 C	170 C	175 C	180 C	190 C	200 C	210 C	220 C	225 C	230 C	250 C	275 C
A	Phenolic	4,750	1,480	1,235		670	184	131					
	Alkyd	4,585		1,232									
A	Epoxy		1,440			622		340					
			1,235			585		331					
A	Polyester					740		177			189		
						771		189			126		
A	Silicone					1,000		539			181		
	Alkyd					847		272			101		
C	Phenolic		1,698				398			96			
	Alkyd												
D	Phenolic				1,724		736		336				
	Alkyd				1,739		554		340				
					2,368		1,126						
D	Epoxy				2,781		1,342						
					2,451								
D	Polyester						3,480						
F	Alkyd				2,306		776		483				
	Melamine												
I	Phenolic						1,302			516		173	64
	Alkyd						1,008			240		165	51
							1,302			420		106	
I	Silicone						952			279		146	44

Equation of regression line: Log life = −4.9086+3,645/C+273.

Table VI. Thermal Life Epoxide-Polyester Enameled Magnet Wire, Unvarnished Specimens

Laboratory	Life, Hours 150 C	175 C	200 C	225 C	250 C
B	11,300	1,735	520	155	92
	9,745 (2)*	1,877	481	177	
		1,785	516	192	
C	6,000 (1)	2,427	500	184	
		1,320	342	86	

Equation of regression line: Log life = −7.3157+4,743/C+273.

* Numbers in parentheses are the number of specimens that failed.

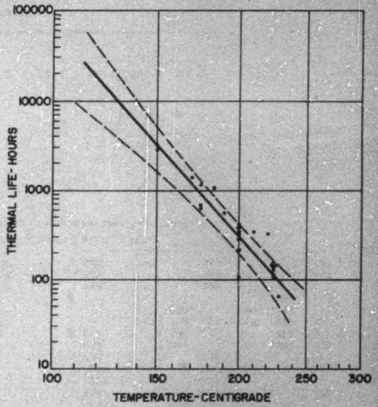


Fig. 3. Thermal life of epoxy enameled magnet wire, unvarnished specimens

Table VII. Thermal Life Epoxide-Polyester Enameled Magnet Wire, Varnished Specimens

Laboratory	Varnish Type	Life, Hours 150 C	165 C	175 C	180 C	185 C	200 C	220 C	225 C	250 C
B	Phenolic-alkyd	7,729 (0)*	7,308 (9)	3,586		1,461	894		127	
B	Organic	7,729 (0)	3,821	2,072		1,064	619		199	
B	Phenolic	9,187	3,307	2,210		847	697		169	
B	Organic	7,729 (0)	6,540	3,560		1,600	743		181	
B	Silicone-alkyd	7,057 (0)	4,412	2,683		1,650	1,008		644	
B	Organic	12,600 (3)	5,112	2,884		1,487	942		261	
C	Phenolic-alkyd	6,000 (0)		1,416			354		55	
F	Alkyd-melamine				2,580		793	360		
I	Silicone						1,008		279	146

Equation of regression line: Log life = −7.0174+4,668/C+273.

* Numbers in parentheses are the number of specimens that failed.

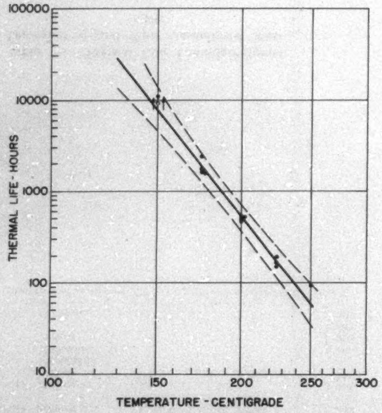


Fig. 4. Thermal life of epoxide-polyester enameled magnet wire, unvarnished specimens

Table VIII. Thermal Life of Polyester Enameled Magnet Wire, Unvarnished Specimens

Laboratory	Life, Hours											
	170 C	175 C	180 C	190 C	200 C	210 C	220 C	225 C	230 C	240 C	250 C	275 C
A	12,000			3,200	1,800	1,100	560		320		140	
						1,175			325		142	
B		14,000			1,400		190	310			63	
					1,558		420	250				
					1,535		460	460				
							480	580				
							520					
							550					
							580					
							190					
							580					
							190					
							580					
							620					
							765					
C		4,000 (0)			2,855			610			29	
					1,176			314				
					1,610			254				
					2,722			322				
E			8,200 (1)*		3,065		1,061		566	221		
			8,200 (1)*		2,660		1,166		567	274		
							1,108		508	229		
							980		480	248		
										241		
G			5,900		1,140			200				
I					4,229			992			148	34
					5,933			520			144	36

Equation of regression line: $\text{Log life} = -9.2682 + 5{,}945/C + 273$.

* Numbers in parentheses are the number of specimens that have failed.

Vol.
Pt. 1
1959

Table IX. Thermal Life of Polyester Enameled Magnet Wire, Varnished Specimens

Laboratory	Varnish Type	Life, Hours												
		150 C	170 C	175 C	180 C	190 C	200 C	210 C	220 C	225 C	230 C	240 C	250 C	275 C
A	Phenolic-alkyd		6,520			1,625		828			287		179	
								738			242		74	
A	Silicone-alkyd							931			487		191	
A	Silicone		12,000			4,690		750			285		169	
			20,000 (0)†			4,950		740			427		85	
			20,000 (3)			4,510		802			615		78	
			20,000 (5)			4,800		1,070			343		83	
			20,000 (1)			4,200		1,490			426			
						3,800		672						
								610						
C	Phenolic-alkyd			9,916			2,820			1,020			262	
							3,650			852			269	
										816			233	
C	Alkyd						3,020			990			255	
C	Epoxy						2,450			528			199	
C	Silicone-alkyd						1,378			638			261	
C	Silicone			21,900			3,360			864			173	
							4,010			546			144	
										682			170	
D	Polyester						6,364		2,206					
D	Silicone-alkyd						3,866							
E	Polyester				>6,200		>6,200		2,362		1,138	491		
					>6,200				2,075		1,063	488		
												558		
												607		
												505		
												703		
												311		
E	Modified phenolic				3,360		2,353				611	298		
F	Alkyd				4,650		2,776		1,802					
							4,584		724					
G	Phenolic	4,500*		1,000*			280*		90*					
G	Alkyd				14,000		2,700			520				
I	Phenolic-alkyd						3,216			1,170			216	
							6,021			1,453				
I	Terephthalic-alkyd						4,368			1,491			511	69
I	Silicone-alkyd						3,476			819			297	114
										712			314	
										650				
I	Silicone						6,370			684			156	71
							4,552			1,045			177	45
										614			104	
										772			156	

Equation of regression line: Log life = $-8.1739+5{,}489/C+273$.

* Failure criterion 2 kv; data not included in calculations of average life.

† Numbers in parentheses are the number of specimens that have failed.

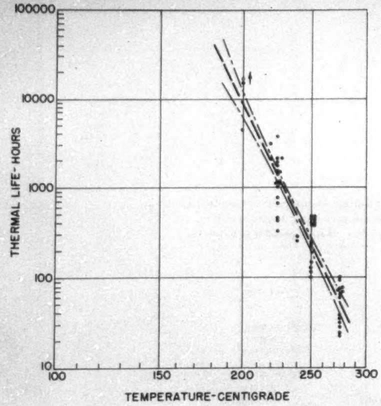


Fig. 5. Thermal life of polyester enameled magnet wire, unvarnished specimens

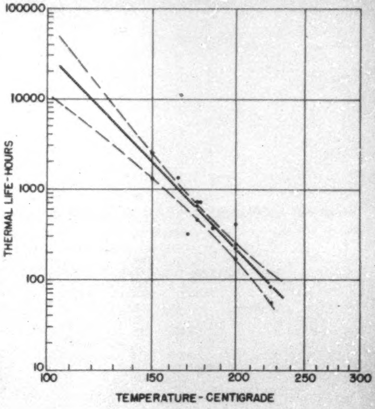


Fig. 6. Thermal life of polyamide enameled magnet wire, unvarnished specimens

Table X. Thermal Life of Polyamide Enameled Magnet Wire, Unvarnished Specimens

	Life, Hours								
Laboratory	150 C	165 C	170 C	175 C	185 C	200 C	225 C	250 C	275 C
A			311						
B	1,300	1,400		470	385	180			
				680		220			
				710					
C	2,755			710		259	58		
I						403	86	82	17.6

Equation of regression line: Log life = −5.6977+3,799/C+273.

Table XI. Thermal Life of Polyamide Enameled Magnet Wire, Varnished Specimens

		Life, Hours						
Laboratory	Varnish Type	170 C	190 C	200 C	210 C	225 C	250 C	275 C
A	Organic-epoxy	349						
		690						
A	Phenolic-alkyd	239	177		93			
I	Phenolic-alkyd			710		168	54	42

Equation of regression line: Log life = −5.6977+3,799/C+273.

Table XII. Thermal Life of Acrylic Enameled Magnet Wire, Unvarnished Specimens

	Life, Hours													
Laboratory	140 C	150 C	160 C	165 C	170 C	175 C	180 C	185 C	190 C	200 C	210 C	220 C	225 C	230 C
A					398				290		62			48
					900				272		47			22
					790				132		72			28
					917				67		72			
											72			
											30			
											27			
B		2,200		1,200		765		350		190		102		
				1,350		518		370		131		59		
D			1,115				584			200				
										206				
E	3,661		976				324							
	2,934		944				422							
	2,802		938				412							
I										279			86	

Equation of regression line: Log life = −7.8690+4,732/C+273.

Table XIII. Thermal Life of Acrylic Enameled Magnet Wire, Varnished Specimens

Laboratory	Varnish Type	Life, Hours: 160 C	170 C	180 C	190 C	200 C	210 C	220 C	225 C	230 C	250 C
A	Phenolic-alkyd		850		244		50			31	19
			845		258		72			60	24
			782		304		72			39	21
							47			95	36
							139				
A	Silicone-alkyd		1,135		413		61				
A	Silicone		644			282	59			39	21
			453			207	48			43	14
							72				
							72				
							43				
D	Phenolic-alkyd			817							
D	Polyester			1,692		480		281			
						419		343			
D	Epoxy							241			
D	Silicone-alkyd	3,877		2,002		1,186		317			
						522					
I	Phenolic-alkyd					506			110		19
I	Silicone-alkyd					730			252		138
I	Silicone					379			101		22
						672			84		57

Equation of regression line: Log life $= -7.4883 + 4{,}677/C + 273$.

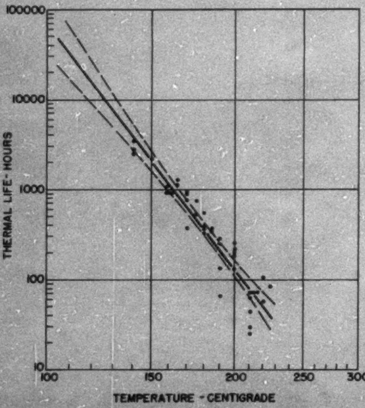


Fig. 7. Thermal life of acrylic enameled magnet wire, unvarnished specimens

Table XIV. Thermal Life of Polyester-Polyamide Enameled Magnet Wire, Unvarnished Specimens

Laboratory	Life, Hours: 150 C	175 C	200 C	220 C	225 C	250 C
B			1,564	369	127	61
C	>6,000	4,399	989		408	
			1,176		314	

Equation of regression line: Log life $= -3.1928 + 2{,}893/C + 273$.

Table XV. Thermal Life of Polyester-Polyamide Enameled Magnet Wire, Varnished Specimens

Laboratory	Varnish Type	Life, Hours: 160 C	175 C	180 C	200 C	220 C	225 C	250 C
C	Phenolic-alkyd		4,300		1,142		432	
			3,147		552		147	
C	Phenolic		3,020		1,411		262	168
C	Epoxy				1,780		655	279
					1,110		403	246
C	Silicone-alkyd				>3,360		1,045	388
F	Alkyd	1,210		706	521	144		

Equation of regression line: Log life $= -3.1928 + 2{,}893/C + 273$.

Table XVI. Thermal Life of Polyurethane Enameled Magnet Wire, Unvarnished Specimens

Laboratory	Life, Hours: 150 C	160 C	170 C	175 C	180 C	190 C	200 C	210 C	220 C	225 C
A			1,050			188		49		
			2,200			44		29		
B	4,546			737			56		70	20
	6,200			1,050			105		85	18
				1,451			190			
				1,627						
				1,417						
C	6,000			401			73			14
D		1,327			510		178			
							203			
							369			
							180			

Table XVII. Thermal Life of Polyurethane Enameled Magnet Wire, Varnished Specimens

Laboratory	Varnish Type	Life, Hours 150 C	170 C	175 C	180 C	190 C	200 C	210 C	220 C	225 C	250 C	275 C
A	Phenolic-alkyd		822			560		84				
			835			238		105				
			1,150			335		90				
			635			98		41				
			576			178		83				
A	Silicone		1,740			130		60				
C	Phenolic-alkyd	6,000 (0)*		518			238			32		
D	Phenolic-alkyd				2,107		680					
D	Epoxy				3,130		1,567		438			
D	Silicone-alkyd				6,746		3,259					
I	Silicone-alkyd						<403			120	19	9
							<403			84	<24	7

Equation of regression line: Log life = $-7.8573+4{,}855/C+273$.

* Numbers in parentheses are the number of specimens that have failed.

Table XVIII. Thermal Life of Silicone Modified Polyester Enameled Magnet Wire, Unvarnished Specimens

Laboratory	Life, Hours 180 C	200 C	220 C	225 C	240 C	250 C	275 C
A				430		125	25
				443			24
B			3,000	760		153	
				1,200		100	
				681			
C		4,770		323			
I		16,377		1,685		196	29
		11,162		1,058		127	32
		17,000 (0)*		3,832		251	71
		17,000 (0)		1,646		398	72
		17,000 (0)		1,831		436	93
		17,000 (0)		2,150		413	79
		17,000 (0)		1,214		228	35
		17,000 (0)		1,726		230	39
		17,000 (0)		1,980		496	100
				1,615		432	60
				2,019		465	64
				1,387		408	68
				1,293		386	71
F	1,384 (0)	1,384 (0)	2,200 (8)		315		
	1,384 (0)	1,384 (0)	2,064		300		

Equation of regression line: Log life = $-12.1727+7{,}616/C+273$.

* Numbers in parentheses are the number of specimens that have failed.

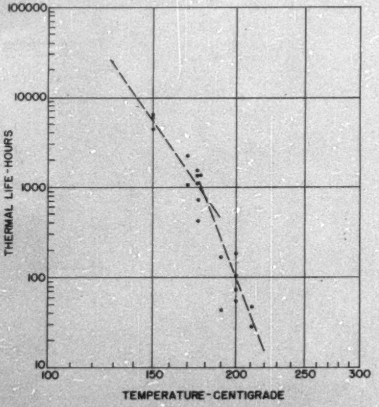


Fig. 8. Thermal life of polyurethane enameled magnet wire, unvarnished specimens

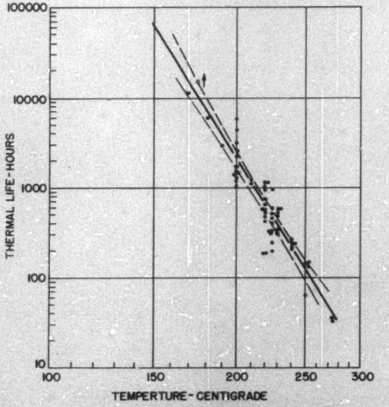


Fig. 9. Thermal life of polyester-silicone enameled magnet wire, unvarnished specimens

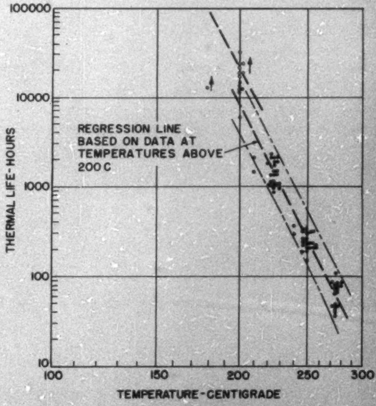


Fig. 10. Thermal life of polyester-silicone enameled magnet wire, varnished specimens

Table XIX. Thermal Life of Silicone Modified Polyester Enameled Magnet Wire, Varnished Specimens

Laboratory	Varnish Type	180 C	200 C	210 C	220 C	225 C	230 C	240 C	250 C	275 C	300 C
A	Silicone			1,580		860	948		253	48	
				2,025			1,100		204		
									309		
C	Silicone		16,800			1,232			273		
						3,145			352		
C	Phenolic-alkyd		1,915*			1,356*			436*		
F	Silicone-alkyd		3,184*		1,679*			584*			
F	Silicone	13,320 (0)	6,195		1,072			376			
					1,856			314			
I	Silicone		31,457 (8)†			1,971			228	76	18
			16,922			1,843			257	74	14
			19,790			2,001			242	72	15
			12,377			1,293			199	44	
			13,116			1,140			217	39	
			13,824			954			214	40	
			14,810			1,259			251	45	
			25,338 (4)			1,702			334	82	
			25,338 (7)			1,678			307	82	
			23,652 (3)			1,612			319	87	
			19,138 (9)			1,212			223	67	
			18,466 (6)			1,565			170	47	
			18,466 (2)			2,035			203	47	
			18,466 (3)			2,207			237	69	
			17,132 (1)			2,517			239	38	
			13,000			1,450			193	76	
			14,548 (0)			2,402			303	84	
			13,757 (0)			2,318			328	87	
			13,104 (0)			1,973			295	119	
			12,432 (0)			2,150			288		

Equation of regression line: Log life $= -11.6181 + 7{,}349/C + 273$.

* Not included in average life calculations.

† Numbers in parentheses are the number of specimens that have failed.

Table XX. Thermal Life of Polytetrafluoroethylene Enameled Magnet Wire

Laboratory	Varnish Type	200 C	225 C	250 C	275 C	300 C
			Unvarnished Specimens			
I		20,000 (0)	21,405 (0)*	257† 5,508	49.2† 708	9.0† 128
			Varnished Specimens			
I	Silicone	20,000 (0)	21,405 (0)	766† 17,776 (2)	152† 1,010	14.7† 182

* Numbers in parentheses are the number of specimens that have failed.

† 2.0 kv test voltage used.

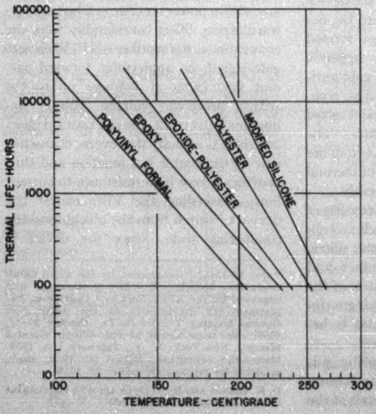


Fig. 11. Relative thermal life of enameled magnet wire, unvarnished specimens

polyester type enameled magnet wire are given in Tables VIII and IX. In Fig. 5 the life of the unvarnished enameled wire specimens are shown.

Polyamide Enameled Wire. Data on the thermal life of one polyamide type enamel are given in Tables X and XI. These data are shown graphically in Fig. 6.

Acrylic Enameled Wire. Thermal life values for acrylic enamels are given in Tables XII and XIII. An indication of the relative thermal life for unvarnished specimens is shown in Fig. 7.

Polyester-Polyamide Enameled Wire. Aging data on a polyester-polyamide-type enamel are given in Tables XIV and XV.

Polyurethane Enameled Wire. Thermal life values for unvarnished and varnished specimens of polyurethane enamels are given in Tables XVI and XVII. Life-temperature characteristics are shown in Fig. 8.

It is apparent in Fig. 8 that the life values for unvarnished specimens do not define a straight line. This fact has been reported by several investigators. Furthermore, the irregularity was noted in the statistical analysis and it was indicated by a low value for the coefficient of multiple correlation. This illustrates one of the reasons for determining thermal life values over wide temperature ranges recommended by this test method if the aging temperatures are below the transition temperature.

Silicone Modified Polyester Enameled Wire. Thermal life values obtained by the Working Group on unvarnished specimens of silicone modified enamels are listed in Table XVIII. Data on varnished specimens are given in Table XIX. The thermal life values for unvarnished specimens are shown in Fig. 9. Two different silicone modified enamels are reported. One enamel that had a relatively poor thermal life when tested without varnish is no longer available commercially.

The difference in thermal life between the two silicone modified enamels is not evident when silicone varnished specimens are tested. The average life of silicone varnished specimens is shown in Fig. 10. Since many sets of samples aging at 200 C are still under test, a life at this temperature cannot be established. Therefore, life values already obtained at 200 C are not included in calculating the average life characteristic.

It is evident that the life values at 200 C do not fall on the line defined by data obtained at higher temperatures. Apparently two deteriorating reactions are taking place. Loss of adhesion of the enamel to copper and chemical reactions between the enamel and copper at high temperatures could be explanations for these phenomena.

These data illustrate why it is so important to have long-time aging tests at temperatures approaching service values before estimating the temperature classification of materials.

Polytetrafluoroethylene Enameled Wire. Only one member of the Working Group has reported data on polytetrafluoroethylene enameled magnet wire. These data, given in Table XX, are included only to give an indication of the relative life that may be expected from this heat-stable coating.

It is also interesting to note the significant decrease in life when a 2.0-kv test voltage is used. Generally, the difference in thermal life obtained using different test voltages is not so pronounced.

Summary

Data from individual graphs are summarized in Fig. 11. These curves, comparing the relative thermal life characteristic for unvarnished specimens, indicate the range of thermal stability among the various types of enamels.

In most cases, more data are needed to establish thermal life characteristic with a greater degree of confidence. Especially needed are life values at lower aging temperatures However, more data will not change the relative positions of the thermal life lines shown in Fig. 11.

Since varnish influences the life of enameled magnet wire, the relative life of varnished specimens cannot always be compared. A study of varnished specimens is further complicated by the fact that a particular varnish may increase the life of one type of enamel and decrease the life of some other enamel. It is suggested that effects of varnish on the thermal life be determined by comparing the life of varnished and unvarnished specimens. In this way, an indication of the degree of increased or decreased life can be estimated.

Evidence given in Figs. 8 and 10 shows that the slope of the life–temperature line may not always be constant. Oxidation of the copper wire, chemical reactions between copper and the enamel, and varying chemical reaction rates may account for this phenomenon. While this phenomenon may occur only for some enamels or for tests conducted at very high temperatures, it does indicate that predictions of thermal life based upon short-time aging tests should be viewed with suspicion.

References

1. A Method for Evaluation of Thermal Stability of Enameled Magnet Wire, C. G. Currin, J. F. Dexter. *AIEE Transactions*, vol. 74, pt. I, May 1955, pp. 227–32.

2. Test Procedure for Evaluation of the Thermal Stability of Enamelled Wire. *AIEE Publication no. 57*, Oct. 1955.

A Silicon-Controlled Rectifier

I—Characteristics and Ratings

D. K. BISSON
ASSOCIATE MEMBER AIEE

R. F. DYER
NONMEMBER AIEE

THE NAME silicon-controlled rectifier has been given to a new 3-terminal semiconductor device which has electrical characteristics similar to a thyratron. This device has been developed by the Semiconductor Products Department of the General Electric Company and has been given the developmental designation *ZJ-39A*. Its characteristics are made possible through the use of a *p-n-p-n* junction configuration. It is capable of handling load currents up to 16 amperes average at peak voltages up to 300 volts. Its fast switching action, small size, and relatively high-current ratings make this new device adaptable to a great many applications. Its characteristics and ratings, which are of utmost importance to the circuit designer, are the subject of this discussion.

Construction

Fig. 1 shows the physical appearance and relative size of the silicon-controlled rectifier. A stud-mounted design has been chosen to facilitate mounting to a heat sink and the top terminal lends itself to either bolt-on or soldered-lead connections.

The internal construction of the device is shown in Fig. 2. The heart of the device is the *p-n-p* silicon pellet formed by diffusion techniques. The upper *n* region is made by alloying the gold antimony into the *p*-type silicon. This completes the *p-n-p-n* structure and gives the device its basic characteristics. The molybdenum disks on top and bottom are included to minimize the effects of thermal expansion on the complete assembly.

The gate or control electrode consists of the small aluminum wire which is connected to the top *p* layer of the silicon pellet. This is an ohmic contact and not a rectifying junction. Control of the device is accomplished by applying a signal to the top *p-n* junction; that is, between the gate and cathode leads.

The subassembly as shown in Fig. 2 is then housed by welding to the bottom molybdenum disk. When the ends of the terminal tubulations have also been welded, a hermetically sealed unit exists. This housed subassembly is then carefully soldered to the copper stud, thus providing a low thermal resistance from the junctions of the device to the external heat sink.

Electrical Characteristics

Basic Electrical Characteristics

Fig. 3 shows the basic voltage-current (E-I) characteristics of the controlled rectifier. The reverse region is typical of any silicon power rectifier cell but the forward region differs considerably from the conventional *p-n* rectifier cell. With zero gate signal, no appreciable forward current flow exists at voltages up to the critical breakover voltage (VBO). Below this VBO point, the cell forward characteristic is actually the reverse characteristic of the center *p-n* junction and thus provides a very high resistance to current flow. Exceeding the VBO causes the device to switch from the blocking to the conducting state. Once the device is

---

Paper 58-1248, recommended by the AIEE Semiconductor Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958, and re-presented at the Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted August 20, 1958; made available for printing December 22, 1958.

D. K. Bisson and R. F. Dyer are with the General Electric Company, Clyde, N. Y.

Vol.
Pt. 1
1959

Fig. 1. General Electric ZJ-39A controlled rectifier

in the conducting state the forward characteristic is typical of any medium-power silicon rectifier exhibiting low dynamic resistance to current flow.

The foregoing explanation of the forward characteristic was based on 2-terminal operation of the device. When operated under these conditions the device will return to the blocking or nonconducting state if the load current is reduced below the value designated as the holding current I_h on Fig. 3. When 3-terminal operation is employed, and sufficient gate current is fed into the gate terminal, the device will then exhibit a forward characteristic similar to that of a *p-n* junction; that is, the forward blocking region will disappear completely.

The theory of operation of the device is beyond the scope of this paper. However, papers have been presented which describe the theory of operation of *p-n-p-n* switches in detail.[1,2]

REVERSE E-I CHARACTERISTICS

Typical reverse characteristics for 125 C (degrees centigrade) junction temperature are shown in Fig. 4. The curves shown are for the various voltage grades to which the cell is now manufactured. The rectifiers are segregated into these several classifications according to their reverse breakdown voltage and the magnitude of the reverse-leakage current at a particular reverse voltage. This PIV (peak inverse voltage) rating procedure is identical to that used on many other types of silicon power rectifiers.

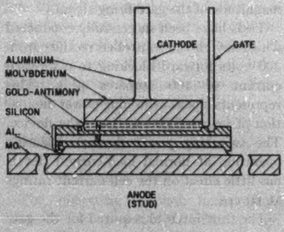


Fig. 2. Internal construction

Data for reverse characteristics such as shown in Fig. 4 have been generated by laboratory measurement since it is difficult to predict mathematically what the reverse leakage currents will be. This is particularly true for the lower voltage grades.

At junction temperatures in the vicinity of the rated junction temperature (125 C) the reverse current at a given voltage increases exponentially with increasing temperature and can be approximated by the equation:

$$I_{RT_1} = I_{RT_2} e^{r(T_1 - T_2)} \qquad (1)$$

where

$T_2 - T_1$ = the temperature rise being considered

I_{RT_1} = the reverse current at T_1

I_{RT_2} = the reverse current at T_2

r = constant of temperature variation -0.06 per C

Equation 1 gives reasonably realistic results when used for the higher voltage grade cells over the temperature range of 125 C ± 50 C. For the lower voltage grades or for lower temperatures the reverse current varies considerably less with temperature than would be predicted by the foregoing equation. This effect is primarily caused by a relatively large surface leakage component of total reverse current which varies only slightly with temperature. The variation of reverse current with temperature is of particular importance in the rating of the device as will be discussed later.

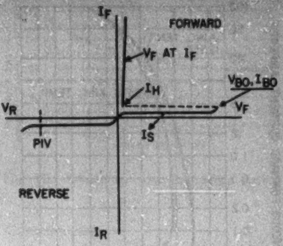


Fig. 3. Basic electrical characteristics

FORWARD E-I CHARACTERISTICS

As previously discussed, the forward E-I characteristics of the controlled rectifier represent two distinct states; the blocking or nonconducting state and the conducting or low-impedance state. Therefore it is more convenient to discuss the characteristics of each state separately.

The forward blocking characteristic is actually the reverse characteristic of the center *p-n* junction and therefore behaves in a similar manner. By simply relabeling the axis on Fig. 4 to read "instantaneous forward voltage" and "instantaneous forward current," the curves become typical for the forward blocking characteristics for the several voltage grades shown. The variation in the magnitude of forward leakage current with temperature is identical to that of the reverse characteristic.

There is one important difference between the reverse characteristic and the forward blocking characteristic of the controlled rectifier. If the maximum allowable PIV rating of the device is sufficiently exceeded, the device will eventually break down and will be destroyed by the avalanche-type currents

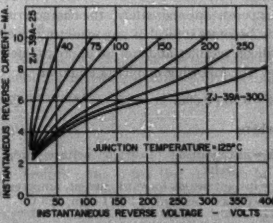


Fig. 4. Typical reverse characteristics

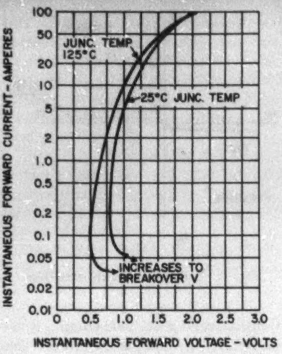


Fig. 5. Typical forward characteristics; conducting state

that would result. In contrast to this, exceeding the VBO of the forward characteristic is nondestructive in nature and will only cause the device to switch from the blocking to the conducting state.

The effects of temperature on the critical value of VBO are of interest. Not all cells follow the same pattern; that is, the majority of present production cells show a considerable decrease in VBO as the temperature is increased, but for others there is little change. Cells are classified according to the lowest value of VBO encountered over the operating temperature range and consequently the cell's VBO rating will apply over the entire operating temperature range.

The forward E-I relationships for the conducting state are shown in Fig. 5. At currents below approximately 100 milliamperes, the forward voltage drop actually increases slightly as the current decreases. This action represents the beginning of the transition between the conducting and blocking states. When the current is decreased below the critical current I_h required to maintain conduction, the device will cease conducting completely and return to the blocking state. At currents above 100 milliamperes (approximately) the characteristic becomes typical of any medium-power silicon rectifier and can be approximated by the following empirical equation:

$$V=0.78+0.0957 \ln i+iR \qquad (2)$$

where

V = instantaneous forward voltage drop, volts
i = instantaneous forward current, amperes
R = device constant of 0.011 ohm, 125 C

Evaluation of the constants is easily made from laboratory data. Equation 2 gives values of forward voltage drop which agree very well with measured values over the operating current range and therefore provides a convenient tool for calculating power generation curves.

The resistance in the last term of the equation has a positive temperature dependence in contrast to a negative temperature dependence for the second term. This causes the curves in Fig. 5 to cross over as shown.

FIRING CHARACTERISTICS

Since firing must be accomplished over a wide temperature range and for a considerable spread in firing requirements from unit to unit, a simple scheme had to be devised to present the information necessary for the design of firing circuits. This effort resulted in the firing characteristic shown in Fig. 6, where the complete range of cell-firing requirements are depicted by means of the shaded area. The nonlinear boundaries of the shaded area represent the maximum and minimum gate E-I characteristics that exist. These E-I characteristics are actually the forward diode characteristics of the top *p-n* junction, the junction between the gate and cathode leads of the cell. Factory testing is designed to assure that all cells will fire within the shaded area. The curve also places maximum allowable gate voltage and current limits of the cells. The maximum limits are primarily dependent on the very small internal gate lead used in the construction of the device. Load lines may be drawn, as are commonly used to depict ignitor firing requirements for ignitrons, which give impedance information necessary for the design of satisfactory firing circuits.

The reverse E-I characteristic of the gate diode (not illustrated) has not been designed to offer high resistance to current flow and consequently a diode in series with the gate is recommended to limit reverse gate current in applications where such conditions exist.

Firing requirements decrease only slightly with increasing anode to cathode voltage over the rated operating voltage range of the device. Cells will operate at applied anode to cathode voltages down to a few volts.

SWITCHING SPEEDS

The time to switch the controlled rectifier from the blocking to conducting state is approximately 1 μsec (microsecond) for typical cells and approximately 5 μsec for the slowest cells. This time includes a short delay time as well as the actual switching time. The switching action is depicted in Fig. 7. The delay time is approximately 0.5–1.0 μsec and is an inherent characteristic of the device. It can be minimized by proper control of the pulse width and magnitude of the gate-firing signal.

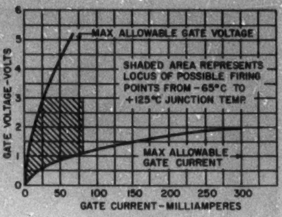


Fig. 6. Firing characteristics

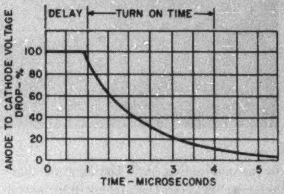


Fig. 7. Typical turn-on time

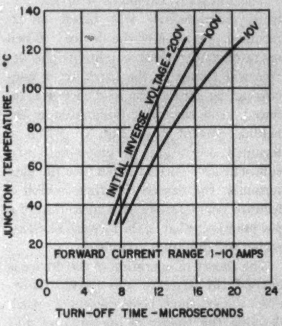


Fig. 8. Typical turn-off time

Tests have been successfully conducted which switch the controlled rectifier from 300 volts forward blocking to a forward current of 100 amperes peak. This represents a peak switching power dissipation of nearly 7,500 watts in the device. The switching power dissipated during operation at normal load-current levels has little effect on the cell-current ratings at 60 cps.

The time interval required for the gate to regain control after a forward current

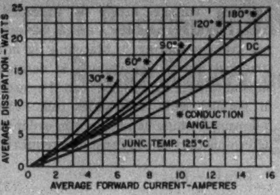


Fig. 9. Forward power dissipation

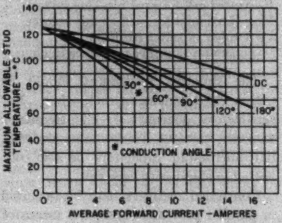


Fig. 10. Maximum allowable stud temperature

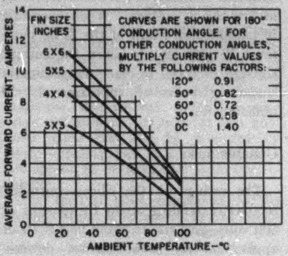


Fig. 11. Required fin size

is interrupted is defined as the turn-off time. The time to turn the device off is somewhat longer than is necessary for turn-on. Fig. 8 shows the relationships between turn-off time, junction temperature, and inverse voltage before turn-off for a typical unit. Forward current before turn-off has not been included as a variable because over the range of 1–10 amperes, for which data were taken, it affects the turn-off time only slightly. Increased forward current causes a slight increase in turn-off time.

Ratings

JUNCTION TEMPERATURE

The operating junction temperature range is −65 to +125 C. The storage temperature range is −65 to +150 C. A low temperature limit is required primarily in order to limit stress in the silicon crystal to safe values. This stress is due to bimetallic bending as the temperature is reduced below the freezing point of the materials used in fabricating the assembly. The upper operating temperature limit was imposed because of the temperature dependence of VBO, the maximum forward blocking voltage that can be withstood. The storage temperature limit of 150 C was selected to achieve maximum reliability and stability of characteristics with time.

THERMAL IMPEDANCE

The internal thermal impedance of the rectifier is defined as degrees centigrade per watt temperature rise of the junction above the cooling stud. The steady state value is 2 C/watt. Transiently, the thermal impedance appears smaller because of thermal storage. Because of the small thermal mass of the silicon pellet, substantial temperature fluctuations appear within a cycle when the supply frequency is 60 cycles. Because of this fluctuation, the rating curves are based on a peak junction temperature of 125 C for 60-cycle circuits. Since there are three junctions in the rectifier, a composite temperature as indicated by the change in forward voltage drop with temperature is used as a means of defining junction temperature.

POWER LOSS

Forward conduction losses are the major source of junction heating. Other losses are forward and reverse blocking loss, gate and switching power losses. Fig. 9 gives conduction loss in watts as a function of average current in amperes for various conduction angles. These curves are based on a current waveform which is the remainder of a half-sine wave which results when delayed angle firing is used. The length of each curve represents the permissible range of current for that particular firing angle. The power curves are the integrated product of the current waveform equation and the forward drop equation; see equation 2.

The maximum allowable stud temperature is shown in Fig. 10 as a function of average current. Conduction, gate, and blocking losses are included and switching loss is neglected. This curve is calculated for an operating frequency of 60 cycles. It becomes conservative at higher frequencies because the junction temperature excursion within a cycle becomes less. At some higher frequency, the heating effect of switching power loss will overcome this advantage.

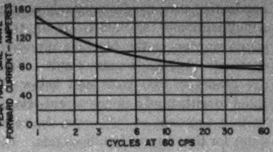


Fig. 12. Maximum allowable surge current

COOLING FINS

In order to maintain the junction temperature within the limits specified by Fig. 10, it is necessary to provide cooling for the stud, i.e., a cooling fin. Fig. 11 gives the required fin size as a function of average current and conduction angle for the following set of conditions:

Fin material	copper 1/16 inch thick
Surface emissivity	0.9 or higher
Air pressure	14.7 pounds/inch^2
Air velocity	free convection
Fin spacing	0.75 inch or more

For other conditions, the stud temperature, as measured by a thermocouple, can be used to determine if the cooling is adequate. For a given fin size, the allowable average current decreases with conduction angle since both blocking and conduction losses increase.

THERMAL STABILITY

The smallest fin size shown on Fig. 11 is 3 by 3 inches. For the cooling conditions given and with a junction temperature of 125 C and inverse voltage of 300 volts, this is the smallest fin size that provides a margin of safety against thermal runaway. Thermal runaway occurs when a critical value of reverse blocking power loss is exceeded. This value of power can be calculated from the following equation:

$$P_{R\,\max} = \frac{1}{Kr} \text{ watts} \qquad (3)$$

where

$P_{R\,\max}$ = maximum stable reverse-blocking loss
K = steady-state thermal resistance from junction to ambient in C/watt
$r = 0.06$

Reverse power loss may be decreased by operating at reduced voltages or junction temperatures.

TRANSIENTS

Voltage transients can be quite severe in circuits using controlled rectifiers because of the relatively fast switching speeds involved. If these transients appear across the rectifier in the forward direction, the rectifier may switch to the conducting state nondestructively causing a malfunction of the circuit. If the transients appear in the blocking direction, destructive breakdown may result. Voltage grades of 200 volts and below are rated to withstand a nonrecurrent reverse voltage transient of 50% over the rated value for times up to 5 milliseconds. For higher voltage cells, the transient-voltage rating is 100 volts above the rated PIV value. For repetitive transients, the peak value should not exceed the continuous PIV rating of the cell.

Forward current transients such as encountered during fault conditions can be withstood up to the values shown in Fig. 12. This curve relates the maximum allowable current peak of a 60-cycle half sine wave to the number of cycles. Under these conditions, the rectifier turns on at the beginning of the current pulse and the turn-on energy dissipated in the rectifier is low. If the rectifier is turned on into a fault after the voltage has built up across the rectifier in the forward direction, the energy dissipated in the rectifier during turn-on can be substantial. Because of this, the current cannot be allowed to exceed 100 amperes until after the turn-on interval has been completed.

When reverse voltage is applied to the rectifier immediately following forward conduction, the rectifier has a low reverse impedance until the stored carriers in the end junctions are swept out. This reverse recovery time varies with forward current, initial inverse voltage and circuit impedance but is normally less than 1 μsec. The maximum allowable current during this interval has not yet been established.

FREQUENCY RESPONSE

The maximum operating frequency for the controlled rectifier is influenced greatly by the circuit application and no single figure can be given. Operating frequency is limited by the turn-off time and switching power losses.

If the time between the end of forward current conduction and reapplication of forward blocking voltage is less than the turn-off time, the rectifier will not have regained control and the circuit will malfunction. The required turn-off time can be estimated from Fig. 8. For example, if a 3-phase bridge connected to an inductive load is phased back 90 degrees to obtain zero output voltage, only 90 electrical degrees are available for the rectifier to regain control. If the turn-off time were 20 μsec, the maximum operating frequency would be limited to 12.5 kc.

In inverter circuits, if the margin angle is small, the frequency limit becomes much lower. As the margin angle is reduced, the initial inverse voltage also becomes less which has the effect of increasing turn-off time. If the margin angle of an inverter were 10 electrical degrees and the turn-off time was 20 μsec, maximum frequency would be 1.39 kc.

Forward switching power loss contributes to average heating and the average current must be reduced as frequency is increased in order to maintain a constant power loss. If the L/R (inductance-reactance) time constant of the circuit is appreciably longer than the turn-on time of the rectifier, the forward switching power loss will be negligible. However, if the L/R time constant is short so that the build-up of current is determined by the collapse of voltage across the rectifier, considerable energy can be dissipated in the rectifier. This energy can be predicted by the following formula:

$$\text{Energy} = 0.65 \text{ by } 10^{-6}\, V_m I_m \text{ watt seconds} \tag{4}$$

where

V_m = anode voltage immediately prior to switching
I_m = anode current immediately after switching

The following example illustrates a calculation of turn-on power:

Circuit	3-phase bridge inductive load
Frequency	400 cycles
Load current	20 amperes
V_m	200 volts
I_m	20 amperes
I_{avg}/cell	6.7 amperes

The average power due to turn-on power would be:

$$P = 400(200)(20)(0.65)(10^{-6}) = 1.04 \text{ watts}$$

This means that for a given fin size as shown in Fig. 11, the average current would have to be reduced by the amount necessary to reduce the average power loss by 1 watt. This can be determined from Fig. 9.

In a similar manner, switching power loss at the end of the conduction interval also requires derating of forward current. At the present time, not enough information is available to predict the energy dissipated during the turn-off interval but it should not exceed turn-on energy because tests indicate that the time required for the rectifier to regain its high impedance state in the reverse direction is less than the turn-on time.

SERIES AND PARALLEL OPERATION

Since the controlled rectifier will turn on without a gate signal if the forward blocking voltage rating is exceeded, series operation requires a means of forcing voltage sharing. Matching is not practical unless the voltage is derated substantially. A resistor can be used in parallel with each series cell. The following formula can be used in order to determine the value of resistance required.

$$r = \frac{30(nV_R - V_{pk})}{n-1} \tag{5}$$

where

V_R = rated cell forward blocking voltage
V_{pk} = maximum forward circuit voltage
n = number of series rectifiers
r = maximum shunting resistance in ohms

The shunting resistances should be within 2% tolerance of each other. For satisfactory sharing of high-frequency transient voltages, it is recommended that 0.01-microfarad capacitors be also connected in shunt with each cell.

Controlled rectifiers may be operated in parallel provided that some means of current sharing is provided. This can be accomplished by external circuit means or by using factory-matched cells. Matched cells are also selected for firing at low anode voltage. When operating in parallel, the width of the firing pulse must be sufficient to allow the current through each cell to reach the minimum sustaining value of approximately 50 milliamperes.

Conclusions

The ratings of the silicon-controlled rectifier, extending up to 16 amperes at 300 volts, give rise to many new fields of application for semiconductor devices. The replacement of thyratrons, magnetic amplifiers, and relays in such applications as static switching, motor speed control, inverters, and voltage surge suppression are a few of the many possible applications.

While far from being complete, the characteristics and ratings have been presented to assist the circuit designer in this work and as a guide for the correct and successful application of this new device.

References

1. PNPN TRANSISTOR SWITCHES, J. L. Moll, M. Tanenbaum, J. M. Goldey, N. Holonyak. *Proceedings*, Institute of Radio Engineers, New York, N.Y., vol. 44, Sept. 1956, pp. 1174–82.

2. MULTITERMINAL PNPN SWITCHES, R. W. Aldrich, N. Holonyak. *Ibid.*, vol. 46, June 1958, pp. 1236–39.

3. THE RATING AND APPLICATION OF A SILICON POWER RECTIFIER, D. K. Bisson. "Rectifiers in Industry," *AIEE Special Publication T-93*, June 1957, pp. 135–42.

CCITT Recommendations for Multichannel Radio Relays and White Noise

CHARLES A. PARRY
MEMBER AIEE

Synopsis: The recommendations of the International Telegraph Consultative Committee (CCIT) and International Radio Consultative Committee (CCIR) now extend into virtually every phase of communications, and are becoming of increasing significance to the radio-relay engineer. Of particular interest is the hypothetical reference circuit proposed as a guide for system design. It is pointed out that recommendations with respect to this reference are intended primarily for international circuits, and do not necessarily apply to all operational conditions. Optimum design would suggest that limits be established on the basis of actual performance requirement.

Because of its significance in radio-relay design, the channel signal-to-noise ratio is considered in some detail, and both nonlinear and thermal components of the noise are discussed. The nonlinear noise is shown to be dependent on both the number of channels and the multichannel signal power (associated with traffic). Distributions for this are given. The effect of relative periods of maximum fading and traffic, and the proportions and periods of maximum international traffic are also taken into account. It is noted that nonlinear noise can also be produced by selective fading, but at present this does not appear to be calculable with any reliability.

The thermal noise component in the channel output is influenced by the amplitude of the received radio-frequency (r-f) signal. During periods of worst fading, on which the design must be predicated, Rayleigh noise distributions may be expected over several links simultaneously, and there is considerable advantage to be gained from diversity operation. It will be shown that the over-all distribution of the multilink system appears to be substantially independent over a fairly wide range, both of the number of fading links and the proportion of these to the nonfading links. This suggests a general design limit for basic system planning.

On this basis it is recommended that the (channel) signal-to-weighted thermal or nonlinear noise ratio should exceed 40 db (decibels) for 99% of the busiest hour for the principal circuits, considering only contributions from the radio equipment. When pre-emphasis and de-emphasis are also taken into account, the basic design requirements for the top channel of an S section system are then: 64.5 db for the signal-to-unweighted thermal noise in an unfaded link; and $(30.5+10 \log S)$ db for signal-to-unweighted nonlinear noise in each section. These figures apply closely for systems from 10 to 40 links in which second-order distortion is predominant.

The use of white-noise testing and its particular relation to these various factors is discussed. It is suggested that the rms power of this signal for an N channel system should be $(-5.0+6.6 \log N)$ dbm0 (db referred to one milliwatt). The signal-to-noise ratio expected in the operational system is readily related to results obtained by this process. The white-noise test is a useful tool under controlled conditions, but measurements in the field may prove unsatisfactory due to the difficulty in separating thermal and nonlinear noise components.

THE INCREASING participation of radio-relay engineers in the planning of national and international telecommunication networks in many parts of the world has emphasized the importance of the work undertaken by the International Telegraph and Telephone Consultative Committee (CCITT).

The recommendations of this group are of special interest, particularly those with respect to the hypothetical reference circuit, while those considerations which relate to per-channel signal-to-noise ratio are of prime concern in system planning.

In order to derive this ratio it is necessary to consider such factors as the distribution of traffic volume, distribution of carrier signal due to propagation, periods of maximum traffic and fading, the proportion of international traffic, and the effects of base-band noise equalization. In addition it is necessary to establish the absolute value for this ratio. One method of determining this is to consider loss of intelligibility and consequent operating time.

It is the purpose of this paper to examine these factors, with a view to establishing a guide in basic system planning. In addition, the use of white-noise testing and the relation of this to required performance parameters is considered.

The discussions are intended to apply to frequency-modulated (FM) systems using frequency division (FD) multiplex.

The CCITT and Recommendations

The International Telephone Consultative Committee (CCIF) and the CCIT were formed officially in 1925 for the purpose of co-ordinating studies being conducted by the various European telephone administrations, and for proposing solutions to the difficulties arising from differences in language and equipment. Since that time they have met regularly, and today all the large administrations in Europe and the Americas and many other parts of the world are members. Although the primary functions remain unchanged, the scope has been broadened to keep pace with the complexity and progress of modern communications.

The CCIR was established in 1927 to study technical and operative radio problems of international interest. At the Atlantic City, N. J., conferences of 1947 this body was reconstituted under the framework of the International Telecommunication Union, and thereby came under the same administrative structure as the CCIF and the CCIT. Recently, the latter two groups were brought together as the CCITT. Some of the responsibilities and functions of these bodies have already been described in the literature.[1,2]

It is the function of the CCITT to think in terms of international operation and to disregard purely national matters, although national aims naturally influence many decisions. National procedures are left to the discretion of the administrations concerned, but many of the studies and recommendations of the CCITT can be applied usefully to purely local networks.

In order to resolve communications problems, it has been the practice to set up groups charged with the responsibility of making specific studies. The reports of these study groups are then brought out as specific recommendations by the parent body concerned, that is the CCIF, CCIT, or CCIR. These recommendations have been improperly referred to on many occasions as "standards." This is not the case. The recommendations reflect the best compromise based on composite and sometimes conflicting views, as well as on the best known state of the art. They also suggest the techniques and operative limits which will permit international circuits to be compatible with each other.

Today, with the rapid expansion of international circuits and American participation in communications planning in many parts of the world, there has come the need for a wider knowledge and under-

---

Paper 59-6, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted February 13, 1958; made available for printing October 16, 1958.

Charles A. Parry is with Page Communications Engineers, Inc., Washington, D. C.

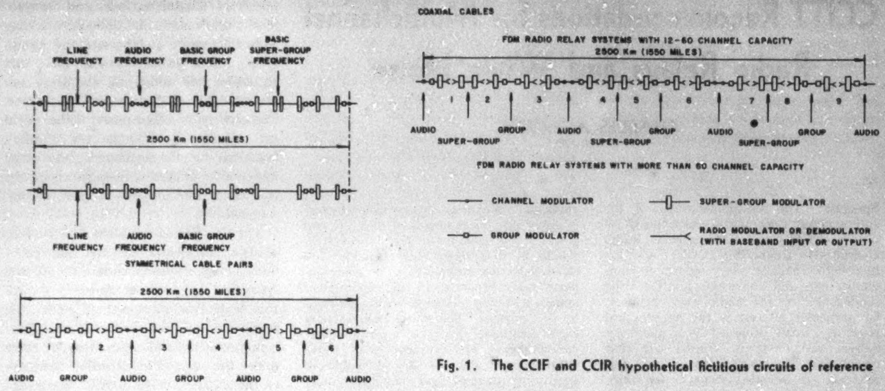

COAXIAL CABLES
LINE FREQUENCY
AUDIO FREQUENCY
BASIC GROUP FREQUENCY
BASIC SUPER-GROUP FREQUENCY
2500 Km (1550 MILES)
LINE FREQUENCY
AUDIO FREQUENCY
BASIC GROUP FREQUENCY
SYMMETRICAL CABLE PAIRS
2500 Km (1550 MILES)
1
2
3
4
5
6
AUDIO
GROUP
AUDIO
GROUP
AUDIO
GROUP
AUDIO
FDM RADIO RELAY SYSTEMS WITH 12-60 CHANNEL CAPACITY
2500 Km (1550 MILES)
1
2
3
4
5
6
7
8
9
AUDIO
SUPER-GROUP
GROUP
AUDIO
SUPER-GROUP
GROUP
AUDIO
SUPER-GROUP
GROUP
AUDIO
FDM RADIO RELAY SYSTEMS WITH MORE THAN 60 CHANNEL CAPACITY
CHANNEL MODULATOR
SUPER-GROUP MODULATOR
GROUP MODULATOR
RADIO MODULATOR OR DEMODULATOR (WITH BASEBAND INPUT OR OUTPUT)

Fig. 1. The CCIF and CCIR hypothetical fictitious circuits of reference

standing of these recommendations. They have always been available in printed form and today all volumes may be obtained in the three principal languages, English, French, and Spanish.[3,4] Containing, as they do, the questions, the reports covering the state of the art, and the recommendations arising from a study of these matters, they are not only a working guide for the engineer, but also an invaluable library of up-to-date condensed information, covering virtually every phase of communications.

In order to evaluate both systems and equipment, the CCIF orginally proposed a hypothetical fictitious reference circuit. Orginally suggested for wire and cable circuits, the concept automatically was accepted for radio-relay planning. The reference circuit is about 1,500 miles long, is divided into a specific number of homogeneous sections, and has a specific number of base-band (audio) patches and frequency translations.[5] The proposed form of these circuits is shown in Fig. 1. For the purpose of analysis they may be considered to consist of 50-microwave links of equal length. This reference circuit concept has the advantage of providing a design framework for both the equipment designer and the system planner. In addition, it is invaluable to the customer as a guide to establishing performance guarantees. The danger in proposing the reference circuit lies in its indiscriminate application to all systems regardless of their actual operational need. It should be recognized that circuits required for international service will demand the highest performance standards. There seems little doubt, however, that somewhat simpler operational requirements exist on routes which do not carry international traffic. Further, the relative number of international circuits which a route has to carry is quite small as can be inferred from Table I.[6] Hence, it is unlikely that there should be any need to impose maximum requirements on all circuits of the international route, nor does it seem reasonable to impose "international" limitations on the "worst" channel in the system. It would seem more in keeping with good engineering principles to apply such limits to the best 5% of channels. It should be evident that different sections and circuits of the national network should have different performance limits in keeping with operational needs. Many private industrial systems are included as part of the national network, e.g., those used by utility companies. Performance standards for these should be based on their own need, not on those applicable to the national network. This has been discussed previously by the author[7] and some suggestions along these lines have been made.

The reference circuit itself is of course only a guide toward detailed specifications for radio-relay systems and for international operation. Reference

Table I. Internal and International Circuits

Country	Total Internal Circuits	Total International Circuits
Argentina	2,714	46
Australia	12,198	
Belgium	15,121	762
Jamaica	73	3
Cuba	507	34
Egypt	2,439	5
France	86,405	1,704
India	3,303	77
Indonesia	362	
Iran	16	
Iraq	560	20
Israel	748	8
Italy	2,376	244
Japan	49,986	29
Malaya	1,535	
Morocco	1,372	59
Fr. W. Africa	304	1
Pakistan	727	25
Panama	71	
Peru	287	
Spain	6,423	91
Switzerland	29,747	866
Syria	338	56
Thailand	13	
Tunisia	781	18
Turkey	609	21
U. S. A.		
Vietnam	28	
Yugoslavia	3,966	122

Note. Circuits include physical phantom and carrier circuits.

Table II. Some Basic Requirements for the International Circuit

Item	Limit
Reference equivalent	≤7 db
Variation of equivalent	±1.7 db
Band of frequencies effectively transmitted*	300–3,400 cycles
Group delay (t)	<150 milliseconds
Phase distortion (from the group delay t)	$t_m - t$ min <10 milliseconds† $tM - t$ min <5 milliseconds
Stability of transmission	±1.7 db
Maximum allowance power at a point of zero relative level	
1. Telephony	not specified
2. Telegraphy	5 milliwatts
3. Photo telegraphy	amplitude modulation 1 milliwatt frequency modulation 0.1 milliwatt
4. Voice-frequency signalling	−26 dbm 0 for continuous tone
Duration of interruption in change-over in stand-by power supply	<150 milliseconds

* The CCIF defines the limiting frequencies as those for which the loss is 8.6 db greater than that at 800 cycles. In the United States, limiting frequencies which are 10 db down with respect to 1,000 cycles.

† m = nominal minimum frequency effectively transmitted. M = nominal maximum frequency effectively transmitted. min = minimum frequency corresponding to minimum group delay time.

should be made to the appropriate volumes of the CCIR and CCIF for this purpose.[5, 8] However, some of the factors which are of prime interest to the radio-relay engineer have been included here. Table II, for instance, lists some of the more important circuit requirements; Fig. 2 shows the over-all frequency response of a channel; Fig. 3 shows the frequency arrangements for the channels of frequency-division multiplex (FDM) systems intended for use over coaxial cables or radio relays; and Table III lists the recommended deviations for FM systems as well as suggested intermediate frequencies.

The Total Noise and Its Proportions

Because the signal-to-noise ratio (as measured at the output of each channel) determines so many basic system-design parameters, it has remained of fundamental importance to the radio-relay engineer. However, the most suitable ratio has never been adequately defined and still remains subject to study and investigation. On the other hand, this should not be taken to imply that satisfactory values for signal-to-noise ratio are not used by the various administrations when designing communication systems. The principal sources of noise which must be taken into account are shown in Fig. 4. The CCIF initially recommended that the noise power in a channel for circuits on coaxial cable should not exceed 10,000 picowatts (micromicrowatts, $\mu\mu$w) at a point −7 db0 (db with respect to a point at zero relative level) under conditions that were somewhat obscure. For open-wire circuits much higher noise levels were permitted. At the Geneva Conference of 1954,[9] a big step forward was made by recommending that the reference signal-to-noise ratio should be based on the long-term mean value. This provides a working point which can be used irrespective of the distributions (in time) with which the noise is usually associated. However, while this is a convenience when adding the noise from various sources, it does not appear to be entirely adequate when it comes to specifying satisfactory performance. At this time also it was considered that noise should be measured during the "busy" hour, and a proposed distribution was given; see Fig. 5. It was also intended as a guide for radio-relay systems. The CCIR[5] in 1956 indicated a general distribution which it considered gave satisfactory telephone service. This is also shown in Fig. 5. Neither of these

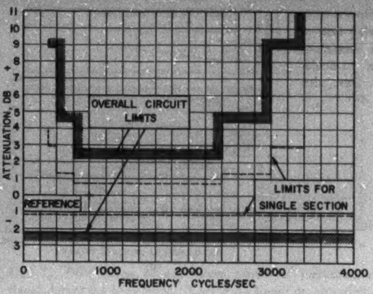


Fig. 2. Frequency response limits for a single channel of the reference circuit. Overall circuit consists of three audio sections in series with necessary limits for each section shown

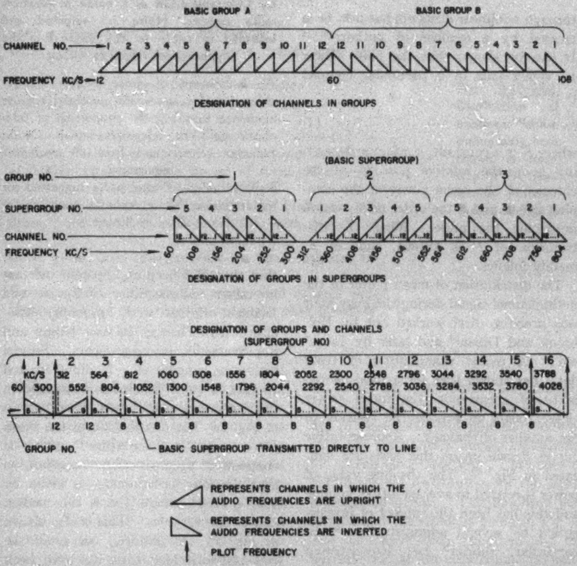

BASIC GROUP A
BASIC GROUP B
CHANNEL NO.
1 2 3 4 5 6 7 8 9 10 11 12 12 11 10 9 8 7 6 5 4 3 2 1
FREQUENCY KC/S
12
60
108
DESIGNATION OF CHANNELS IN GROUPS
(BASIC SUPERGROUP)
GROUP NO.
1
2
3
SUPERGROUP NO.
5 4 3 2 1
1 2 3 4 5
5 4 3 2 1
CHANNEL NO.
FREQUENCY KC/S
60 108 156 204 252 300 312 360 408 456 504 552 564 612 660 708 756 804
DESIGNATION OF GROUPS IN SUPERGROUPS
DESIGNATION OF GROUPS AND CHANNELS
(SUPERGROUP NO)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
KC/S
60 300
312 552
564 804
812 1052
1060 1300
1308 1548
1556 1796
1804 2044
2052 2292
2300 2540
2548 2788
2796 3036
3044 3284
3292 3532
3540 3780
3788 4028
12 12 8 8 8 8 8 8 8 8 8 8 8 8 8
GROUP NO.
BASIC SUPERGROUP TRANSMITTED DIRECTLY TO LINE
REPRESENTS CHANNELS IN WHICH THE AUDIO FREQUENCIES ARE UPRIGHT
REPRESENTS CHANNELS IN WHICH THE AUDIO FREQUENCIES ARE INVERTED
PILOT FREQUENCY

Fig. 3. Arrangement of channels, groups, and supergroups for the coaxial cable and radio-relay systems

Table III. Some Preferred Characteristics of Radio Relays (FM Systems Using FDM)

Frequency Deviation Without Pre-emphasis		Base Band	
Maximum Number of Channels	Rms Deviation per Channel, Kc	Frequency Limits of Base-band, Kc	Nominal Base-band Impedance, Ohms
24	35	12–108	150 balanced
60	50, 100, 200	12–252	150 balanced
		60–300	75 unbalanced
120	50, 100, 200	12–552	150 balanced
		60–552	75 unbalanced
240	200	60–1,052	75 unbalanced
600	200	60–2,540	75 unbalanced

Intermediate Frequencies:
35 mc for radio frequencies of 50 to 1,000 mc
70 mc for radio frequencies above 1,000 mc
Nominal impedance of the intermediate frequency circuits, 75 ohms, where interconnection between systems is to be made.

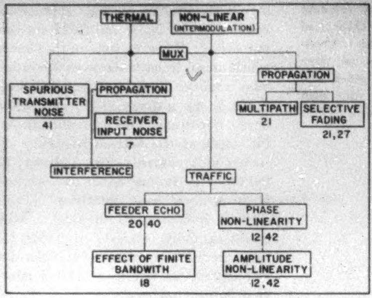


Fig. 4 (left). Principal sources of noise in radio-relay system. Numbers are the references which discuss the particular noise concerned. See also reference 44

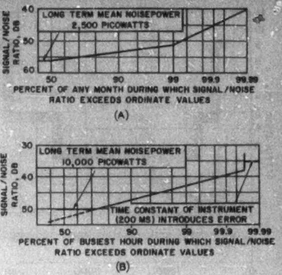


Fig. 5. Tentative noise distributions suggested by CCIF and CCIR as a guide in planning radio systems. Noise is weighted and intended to apply to any channel in the hypothetical reference circuit

A—Distribution of noise power in any channel during any month; unlikely to cause hindrance to substantial proportion of telephone calls or telegraph signals. Noise excludes contributions from the multiplex equipment

B—Distribution of total noise suggested for busiest hour. Noise includes contributions from multiplex

distributions is entirely suitable for the design of radio-relay systems, and at present the only real solution is to aim for a performance agreeable to administrations and consistent with good engineering practice. It is the present intention to examine further this question of the signal-to-noise ratio.

The suggested proportions of mean noise power in the channel of the hypothetical circuit at a point of zero relative level have been given by the CCIR[5] as follows: 1. 2,500 $\mu\mu$w due to the multiplex equipment; 2. 5,000 $\mu\mu$w in the busy hours in the absence of fading; and 3. 7,500 $\mu\mu$w in any hour.

These proportions recognize that periods of maximum traffic and worst fading rarely coincide. This effect is indicated in Fig. 6 and is also discussed later.

Traffic and Nonlinear Noise

The distribution of nonlinear noise (also referred to as intermodulation and distortion noise) among the channels of the multichannel system, due to the transmission of the complex signal through nonlinear elements has now been treated by a number of authors.[10-12] It can be shown that the noise power p_n which falls in a particular channel is given by:

$$p_n = KyP^{\beta} \tag{1}$$

where K is a constant, y a factor depending upon the relative position of the channel, P the mean power of the complex signal, and β the order of distortion producing the noise. When the distribution of P is known, that of p_n readily follows.

The distribution of mean power of the multichannel signal during the busy hour was probably first worked out by Holbrook and Dixon[13] and later by Jacobsen.[14] A recent examination of these data[7] has shown that for a fixed probability of occurrence, the mean power closely bears an exponential relation to the number of channels. Representative curves drawn up on this basis are indicated in Fig. 7. The absolute value of power is related to average talker volume and this has been the subject of investigation by several administrations.[9] In particular, Subrizi[15] has demonstrated that there has been an apparent increase in talker volume since Holbrook and Dixon's original work, probably associated with changes in user habits and improvements in equipments. Subrizi arrives at a figure of −15 dbm0 for toll circuits, a value which has been used in the derivation of the mean power curves of Fig. 7. The distribution of the mean power (with time) for a fixed number of channels is approximately Gaussian, so that the long-term average is given by $Vm+0.115\sigma^2$, where Vm is the median and σ^2 the variance. This is also shown in Fig. 7. In addition, curves of the rms test-tone load capacity have been included. Curve f is that obtained from reference 13, using an overload expectancy (as defined by Holbrook and Dixon) of about 10^{-5}. The effects of limiting on all channels is included. Curve e is that derived by the author[7] using an overload expectancy of 10^{-3} and based on Subrizi's figure for the average talker. An allowance is also made for limiting. However, curves a, c, and d, for the mean power, do not include the effects of limiting and as such, therefore, provide a small design margin. It should also be pointed out that the figure used for average talker volume is still probably 4 to 5 db too great for short-haul circuits[15] so that the mean power curves of Fig. 7 are probably much too conservative when applied to short links.

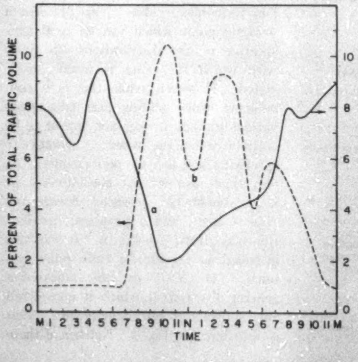


Fig. 6. Curves prove that, under typical conditions, periods of maximum traffic and worst fading do not coincide

Curve a = typical diurnal fading on a microwave link; smoothed data[22]

Curve b = total traffic volume based on averaged data for 25 principal American cities taken from information supplied by the American Telephone and Telegraph Company

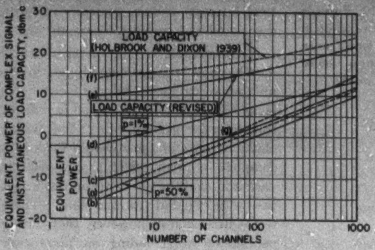


Fig. 7. Distribution of equivalent power of multichannel signal during the busy hour. Curves are based on an activity factor $\tau=0.25$. p is the probability that ordinate will be exceeded

Curve a = median value of the distribution
Curve b = power assumed for signaling, carrier leaks, and pilot tones
Curve c = long-term mean power of distribution
Curve d = power in the signal due to traffic alone, exceeded 1% of the time
Curve e = instantaneous load capacity recommended in reference 7 and based on a peak-to-rms ratio with probability of 0.001. Long-term mean power of single talker −15 dbm0. Includes limiting
Curve f = load capacity recommended by Holbrook and Dixon,[13] and based on peak-to-rms ratio with probability 10^{-5}. Includes limiting
Curve g = rms level of white-noise test signal suggested by CCIR
Curves a through d apply for uncontrolled volumes with effects of limiting not included. Long-term mean power of single talker −15 dbm0

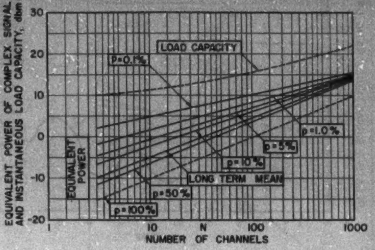


Fig. 8 (right). Curves similar to those of Fig. 7 but recalculated to include effects of the tones of curve b. Suggested for practical long-haul toll circuits

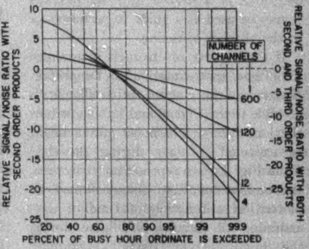


Fig. 9 (right). Distribution of nonlinear noise during busy hour shown as function of the number of channels (based on Fig. 7)

The signal transmitted to line includes carrier leaks and signaling tones, which have not been taken into account in deriving the data for Fig. 7. Some allowance must be made for these signals in order to establish practical design curves for the complete multichannel signal. The CCIF has proposed in-band signaling for international circuits, so that the net effect of signaling tones during the busy hour will be negligible. However, there are systems using out-of-band signaling with "tone on" during speech. The maximum power for a continuous signaling tone as proposed by the CCIF[16] is −26 dbm0, the value which will be assumed for an out-of-band tone. The CCIF also gives −14.5 dbm0 as the limit for the combined effect of carrier leaks in a 12-channel system which, on a power additive basis, averages −24 dbm per channel. The combined effect of leaks and tones is then −22 dbm0 per channel. If some margin is allowed for various discrepancies, a figure of −20 dbm0 per channel would seem reasonable. It is therefore suggested that, in a system with all channels busy, the net effect of all these additional tones may be taken as −20 dbm0 per channel, also shown in Fig. 7. The mean power due to these tones may now be combined with the distribution curves established for speech alone (see Fig. 7) in order to obtain the net distribution for the signal transmitted to line. This has been performed and the result is given in Fig. 8. The data have been smoothed and the nearest straight-line approximations used. The result is a rather too rapid convergence for very large numbers of channels, but it is considered that the error involved is sufficiently small to be neglected in favor of the exponential convenience.

Most designs are based on the $p=1\%$ curve, so that the effect of the additional tones is virtually negligible when the number of channels is small. For large numbers of channels, however, the effect of the tones is more significant, and a substantial part of the power is associated with the spurious signals.

The curves of Fig. 8 give the distribution of P in equation 1. The noise distribution for various orders of distortion is now readily obtained since the noise is proportional to $P\beta$. It is immediately apparent that the noise distribution due to traffic is dependent both on the order of distortion and the number of channels. This is shown in Fig. 9. Since the CCIF has suggested that the reference signal-to-noise ratio should be based on the long-term mean value, the curves of Fig. 9 have been prepared on this basis. In a practical system, it may be expected that $3>\beta>2$, so the ordinate scales of Fig. 9 have been provided for the two conditions: 1. the more usual case when second-order distortion predominates and $\beta=2$; and 2. the general case when both second- and third-order distortion are present, and the net distribution lies about half way between those which are obtained for $\beta=2$ and $\beta=3$. The curves neglect the effects of overload distortion. Noise from this source would tend to increase the slope of the curves, but in practical systems no noticeable effects would be expected for abscissas less than 99%. The influence of bandwidth restriction[17,18] is also disregarded here since the system design usually permits this to be made negligible. It should be recognized that the distributions of Fig. 9 apply to any part or parts of the over-all circuit, only the absolute value being affected by the number of circuits in tandem. There will, however, be a slight change in distribution as the circuit length is extended, since usually third-order effects become more noticeable. However, Fig.

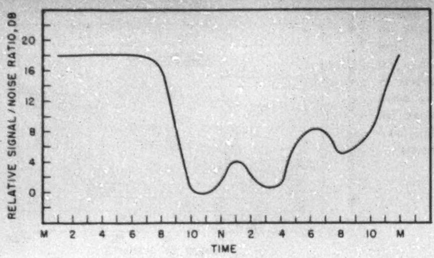


Fig. 10. Possible daily variation in long-term mean nonlinear noise power

9 may be used as a guide to performance in the absence of fading.

If the number of channels is assumed to a first approximation to be proportionate to the volume of traffic, it is possible to use Figs. 6 and 7 to show how the long-term mean noise power might be expected to vary during the day. From Fig. 6, for a representative system, the number of busy channels may be plotted as a function of time of day, assuming all channels are busy during the "busiest" hour. Using this number, the variation in the long-term mean signal power may be obtained with the aid of Fig. 7. The variation of mean noise power is then obtained by applying equation 1. In order to take a representative case, this process was carried out for a 60-channel system, with second-order distortion predominant. The results are in Fig. 10. The residual noise during quiet hours is due to the presence of carrier leak. In deriving the curve of Fig. 10, the carrier leak is taken as −24 dbm0 per channel, and signaling tones as −26 dbm0 per busy channel. This curve probably gives no more than a general picture of the probable diurnal variation, and many departures could be expected in practice. So far, very little information is available on these statistics.[19]

A further factor, the hours of usage, cannot be disregarded. It is extremely unlikely that international traffic would be carried during the periods of maximum local traffic. In the absence of a specified correction factor for this, it is suggested that the loading be based on the number of channels, excluding international circuits. It can be expected in practice that the number of busy channels will never exceed this, since operation of the international circuits will occur at a time when there has been at least a commensurate decrease in the number of local channels in use. This correction is suggested in order to lead to optimal design of radio-relay systems.

Nonlinear Noise Due to Propagation

Periods of deep fading are frequently attributable to multipath effects with consequent distortion and nonlinear noise. Means for calculating this when the specific conditions of fading are known have been derived.[20, 21] However, a detailed study of the phenomenon leads to the conclusion that at the present time it is not possible to derive an expression for this noise from the depth and rate of fading alone.[21] It is quite likely that under certain conditions, noise from this source may be much greater than that associated with either the input thermal noise or with nonlinear noise arising from distortion in the equipment itself. Unfortunately, apart from narrow beams and obstruction-free paths, there seems little which can be done either to predict or protect in this situation. Medhurst gives a simple expression for the maximum distortion due to a deep fade associated with a single echo, assuming a white noise signal. His expression for the noise-to-signal ratio N/S in the top channel of the FM system is

$$N/S = t^2 \times d^2 \times f_d f_m \times 0.14(2\pi)^2 \qquad (2)$$

(voltage ratio)

where f_d is the frequency deviation in cycles, f_m the highest base-band frequency in cycles, t the delay in seconds, and f_d the depth of fade, defined as the ratio of unfaded to faded signal amplitudes. A quick calculation will show that even for an echo of only 1 millimicrosecond and a fade of 30 db, wide-band systems will suffer severely from this cause.

Thermal Noise at Receiver Input

The principal source of thermal noise is that associated with the input to the radio receivers. This means that the per-channel signal-to-thermal noise ratio is directly proportional to the carrier power (at least over the ranges of practical interest) and is, therefore, directly associated with propagation conditions.

A considerable amount of experimental data has now been presented in relation to microwave circuits[22−28] but much of this lacks definition and detail and is difficult to relate the general case. The fading characteristics of the separate links in a complete system exhibit considerable variation, and the net distribution cannot easily be related to one link. There seems little doubt, however, that during periods of worst fading on microwave circuits, a Rayleigh distribution is approximated, and this permits an approach for the worst case. Gudmandsen and Larsen[24] indicate that the distribution on specific circuits was more severe than the Rayleigh for only 7% of the observed days and 25% of the observed hours (maximum figures). Cabessa, Byam and Millar, Halena, and Kaylor, to mention a few, also indicate this tendency toward Rayleigh distribution. Gough, on the other hand, observing the band from about 80–200 megacycles, suggests a Gaussian distribution as suitable for diurnal variations, but his sampling interval (from 1 to 2 minutes) is perhaps too long compared to the sampling period to show the fine grained fading structure.

Fading exhibits both a diurnal characteristic, as shown in Fig. 6, and a seasonal variation. Severe fading usually occurs for only a relatively short period of the year. During less severe fading periods, the distribution tends toward Gaussian, with a standard deviation (diurnal) of less than about 3 db. The nature of the Rayleigh distribution is such that a rather large fading margin must be provided if satisfactory reliability is to be achieved. Consequently, a design based on worst fading periods will be more than adequate for the rest of the time.

The data which are presented relative to Rayleigh fading generally indicate that with this type of fading, the median signal level is depressed several db below the normal unfaded value. Theoretically it can be shown that the long-term mean noise power with Rayleigh fading is approximately 10 db greater than the noise when there is no fading at all,[8] a figure which seems to agree reasonably well with the value indicated in the literature. This leads to the following expression for the relation between the mean noise power N_m at the end of m links in terms of the mean noise power N_o in one link without fading

$$N_m = m N_o \left[10 \frac{n}{m} + 1.44 \frac{m-n}{m} \right] \qquad (3)$$

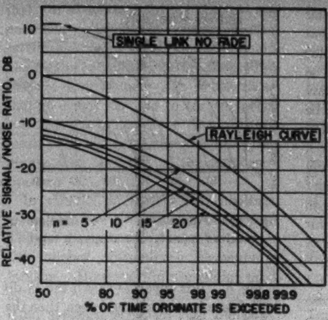
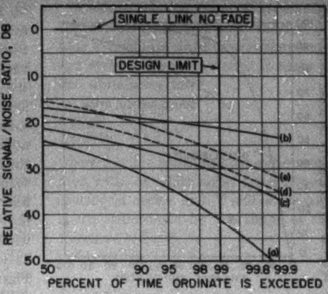


Fig. 11. Distribution of thermal noise in channel due to independent Rayleigh fading on n links and no fading on 50 −n links

Fig. 12 (right). Distribution of thermal noise in channel due to fading and nonfading links. Selector-type diversity is assumed

Curve a = 15 links with Rayleigh fading and 35 without
Curve b = correlated Gaussian fading on 25 links, 25 without
Curve c = same as curve a but assuming dual diversity on all links
Curve d = independent Rayleigh fading on all links of a 10-link system with diversity
Curve e = same as curve d but with 5-link system



where *n* is the number of links with Rayleigh fading and the factor 1.44 is an experimentally observed correction. Distributions based on this relation are shown in Fig. 11. The necessary allowance has been made for the 1.6-db difference between the mean and the median in the Rayleigh distribution.[28]

Severe fading is usually associated with seasonal characteristics, so that it can be expected that this will occur over a relatively long portion of the route. A study of weather conditions over typical 1,500-mile paths indicates that nearly one third of this might be subject to severe conditions at any one time. Accordingly, the reference for severest fading conditions will be based on 15 fading and 25 nonfading links. The net distribution due to independent Rayleigh fading on such a system is shown in Fig. 12, and can be assumed applicable to any of the worst hours.

When the fading range is small, it is usually associated with changes in refractive index which occur over a considerable area. This implies possible correlation of the fading phenomena, which generally exhibit a Gaussian distribution. Accordingly, it will be assumed that during these periods the distribution will be due to 25 links with Gaussian fading, and 25 links with signal about equal to the median value for one link. In the worst case the fading will be correlated and this is shown in Fig. 13.

During severe fading it has been found that virtually the full effect of uncorrelated diversity action can be achieved.[24,25] The theoretical improvement is readily available from the literature.[21] The effect of dual selector-type diversity is also shown in Fig. 12. It is readily seen that a system designed on this basis will be quite satisfactory when correlated Gaussian fading occurs over most of the route.

The net distribution indicated by curve *c*, Fig. 12, is closely Gaussian over the range of interest. It can be seen also from Figs. 11 and 12 that the net distribution does not appear to be substantially affected even for fairly large changes in the number of links, the proportion of fading links, or, when fading is not severe, by the change in the general form of the distribution. This suggests that a generalized design procedure based on curve *c* of Fig. 11 might be used for system planning. The distribution would be only slightly different when a short section consisting of only a few links, is being considered. Of course, the mean value would also change as indicated by equation 3. Using this and Fig. 11, the design limit for a single unfaded link can be easily established.

The Noise Limit

One of the difficulties which is presented to the system engineer once a noise distribution has been determined, is to establish the absolute value for some degree of reliability. The general figure of 50 db for signal-to-noise ratio has been accepted as suitable but its origin is obscure and just whether it is in excess of real requirements has never been satisfactorily determined.[7] The question then is, just what does a specific signal-to-noise distribution mean as far as the listener is concerned? Burrill[29] has made some useful observations in terms of judged listener quality for program transmissions but these data are not entirely applicable to the telephone channel. A considerable amount of information has been obtained on the effect of various signal-to-noise ratios on articulation.[30−34] After studying this literature, it is apparent that satisfactory and intelligible speech may be conducted with extremely high noise-to signal ratios; much higher than is customarily associated with even poor-quality voice channels. When the noise is interrupted, corresponding for instance to the effect of short (less than 1 second) deep fades, the tolerable level of noise is great enough[33] to preclude any consideration of the effect of these fades on intelligibility in practical systems. What, therefore, should be the judgment criteria? It is known that noise will reduce articulation, and with this there is an increase in the time taken to understand a sentence.[34] This leads to the concept that the system might be designed on the basis that the increase in time during a busy hour from this cause should not exceed, for example, 1%, a figure in keeping with other design parameters related to the busy hour. It is interesting to note that Collard[35] suggested time efficiency as a general criterion of transmission performance, but his paper did not appear to receive much further attention. The increase in time is related directly to the decrease in intelligibility and may be obtained from curves of intelligibility and articulation.[34] Fig. 14 shows intelligibility versus sound and syllable articulation.[9,34] The CCIF

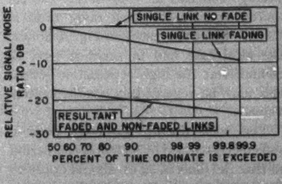


Fig. 13. Distribution of thermal noise in channel due to correlated Gaussian fading. System of 50 links assumed with fading on 25 links simultaneously and no fading on remaining 25 links

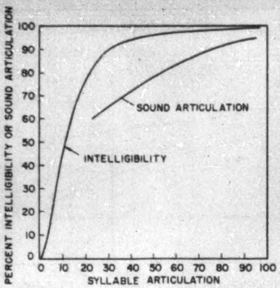


Fig. 14. Sentence intelligibility versus sound and syllable articulation

uses 80% sound articulation as a reference figure which corresponds to about 96% intelligibility on these curves.[1]

Fletcher[31] has determined the articulation for a typical commercial telephone circuit over a 9-db trunk. In addition, he has derived curves showing the effects of white noise received over an ideal circuit and these data may be used for evaluation of noise limits. Fig. 15 is based on this information. The intensity level of received speech when the talker is in the presence of typical room noise will probably be between 70 to 90 db, with maximum articulation above 95%. Fig. 14 shows that this corresponds to almost 100% intelligibility.

It can be seen that the distribution of noise with time will cause a corresponding distribution of intelligibility. If the noise during the busy hour causes a maximum loss of 2% in intelligibility, it can be expected that the average loss would be 1%, corresponding closely to an increase of 1% in operating time. Using Fig. 14, the 2% limit is reached with a syllable articulation of about 68%. Fig. 15 indicates that this is true when the noise is approximately 14 db below speech level. This noise level is sufficiently above room noise[36] for the latter to be negligible. If we assume an average toll call at −15 dbm0, the required signal-to-weighted noise ratio is 32 db. Since the curves of Fig. 15 are for the ideal case and in practice tend to compress at the high articulation levels, this may be regarded as a conservative value.

The noise during the busy hour is sufficiently defined by 99.9% of this period so that the limit of 32 db should apply to the 99.9% abscissas on the distribution curves. It is more convenient, however, to speak in terms of the ratio which is exceeded for 99% of the time. Using Figs. 9 and 12 as a guide, a value of 38 db is suggested. It would also appear from examination of Figs. 14 and 15 that there would be little practical effect for variations in this value of as much as ±5 db. On a similar basis, for private industrial systems where limits might be somewhat relaxed, a signal-to-weighted noise ratio of 33 db exceeded for 95% of the hour should be adequate. A limiting value for any design (30 db for 95% of any hour) seems reasonable. These figures do not consider military applications for which a signal-to-weighted noise ratio of as low as 13 db may be acceptable.[37] Figs. 14 and 15 substantiate the fact that short periods with extremely high noise levels are not likely to affect intelligibility appreciably. It is also apparent that different noise distributions with the same mean value may produce different articulation loss. Thus, on the basis of present reasoning it would appear undesirable to try to specify performance in terms of the long-term mean value alone.

Base-Band Noise Equalization

The practice of using pre-emphasis and de-emphasis networks in FM systems has come about in order to redistribute the noise more evenly over the base band. Reference has been made elsewhere to base-band noise equalization, and a method for the calculation of improvement in top channel noise has been outlined.[38] The CCIR[5] has suggested tentatively that the maximum equalization between the top and bottom ends of the band should be 10 db. This corresponds closely to the suggestion that the maximum insertion loss of the network should be about 16 db.[38] This network will also produce a virtually flat noise spectrum over the base band. The following improvement in the top channel signal-to-noise ratio is obtained: 1. Thermal noise only, 3 to 5 db; 2. Second-order nonlinear noise only, 6 to 5 db; and 3. Equal thermal and nonlinear (second-order) noise, 4 to 5 db.

These figures are for the noise due to the radio equipment alone, and are used in the following paragraphs.

Total Noise Distribution

With an absolute value for the noise determined and with knowledge of the distributions, it is a relatively simple matter to establish fundamental design parameters. In the first place, it may be assumed that the multiplex equipment produces a mean noise power of −56 dbm0, with a distribution probably similar to that laid down by the CCIF and shown in Fig. 5. It is apparent that if the 38 db 99% figure is used as a reference, then the influence of the multiplex is virtually negligible. An increase of 1 db in the ratio for radio equipment alone is adequate to take this latter into account. An additional 1 db is suggested as a margin, resulting in a figure of 40 db for radio equipment alone.

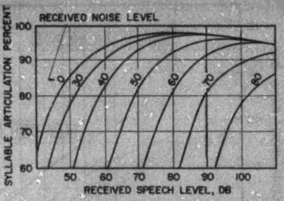


Fig. 15. Articulation as a function of received noise level and speech level

It is difficult to ascertain whether the nonlinear noise distribution of Fig. 9 and the thermal distribution of Fig. 12 should be considered as acting together or separately. The evidence indicates that maximum fading during periods of maximum traffic occurs as an exception, and it is doubtful if a general design recommendation should be established on this basis. Nevertheless there is always some traffic during fading, just as there is always some fading during the busiest hour. Furthermore, it is apparent that the distribution which may apply to the over-all reference circuit will not apply precisely to a small section. During periods of worst fading this again will surely coincide with hours of maximum traffic. However, if the two distributions do occur and the noise powers corresponding to the 1%-of-time points are made equal, there is no appreciable degradation of the system performance and the matter would seem to be largely academic. Indication is shown by the combined distribution in Fig. 16. Thus, it is sufficient

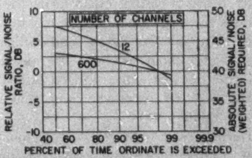


Fig. 16. Combined distribution of thermal and nonlinear noise assuming the separate distributions of Fig. 12, curve (c), and Fig. 9 occur simultaneously. Noise due to each separately, is adjusted to give 40-db signal-to-noise ratio for 99% of the time

that the signal-to-weighted thermal and nonlinear noise ratios considered separately, associated with the radio equipment alone, should be better than 40 db for 99% of any hour. Fig. 16 also shows that it is not suitable to try to specify performance in terms of the long-term mean value alone, because this must then be specified in terms of the number of channels.

In general, it appears that nonlinear noise adds power-wise along the system. That is, the noise power is proportional to the number of nonlinear elements on tandem. However, voltage-adding products may become significant on long wide-band systems, in which case the power is proportional to the square of the number of contributing elements. On the basis of power addition, the signal-to-weighted noise ratio required in the top channel of each link or section of the systems will be $(40+10 \log S)$ db where S is the number of homogeneous sections. The general design requirement, assuming second-order distortion, is then $(30.5+10 \log S)$ db for the signal-to-unweighted nonlinear noise ratio prior to the application of base-band noise equalization. This reference ratio should be established for the multichannel signal given by the 1% curve of Fig. 8.

In the case of thermal noise, it is seen from Fig. 12 that the fading margin per link should be 41 db for a nondiversified system and 31 db with dual diversity, to cater to the worst hours during the year. The latter figure corresponds to a design value in the top channel for signal-to-thermal noise (unweighted) prior to the use of pre-emphasis and de-emphasis of 64.5 db per unfaded link when base-band noise equalization and weighting are taken into account. This requirement holds closely for a wide variation in the number of links and the proportion of fading to nonfading links, but for fewer than ten links new curves should be drawn up in a manner similar to those of Fig. 12. When the operating frequencies are sufficiently low and the path conditions will not produce severe fading, representative curves similar to those of Fig. 13 can be drawn up, e.g., Gough's data.[36] If uncorrelated Gaussian fading is expected new distribution curves may be drawn.[37]

Propagation Outage Time

Thus far no mention has been made of outage time due to deep fades. This has not yet been defined by the CCIR, nor has the matter been covered in the literature. With the FM system, deep fades below the threshold will cause the noise to rise to a very high value. As already mentioned, with the rather high signal-to-noise ratio pertaining during the worst hours, it is extremely unlikely that this noise increase will cause appreciable interference with speech. The sudden loss of signal tends to clip an element of time from the speech sounds. The effect is likely to be more significant in any loss of articulation. In addition, it is customary for the multichannel system to carry telegraph signals and these may be disrupted for the outage period. The acceptable percentages of time for these phenomena have not yet been determined. However, if the frequency and average duration of fades[38] are considered in conjunction with the r-f signal distribution which dictates Figs. 11 and 12, it would seem reasonable to believe that if this period does not exceed 1% of any hour (during conditions of worst fading), the performance would be acceptable. The implication is that the threshold r-f signal should be not less than that corresponding to the 99% abscissas of Fig. 12. It should be mentioned, perhaps, that this criterion probably would apply irrespective of the absolute signal-to-noise ratio pertaining to the threshold carrier.

White-Noise Testing

The principle of white-noise testing has now been well established in the art.[7,39] The CCIR[5] tentatively has suggested that the rms value of the test signal for systems of more than 240 channels should be given by $(-15+10 \log N)$ dbm0, where N is the number of channels (curve g of Fig. 7). It would be more appropriate however, to choose a relation which would fit the 1% curve shown in this figure. The rms signal would be $(-5.0+6.6 \log N)$ dbm0 and is suggested because, as already indicated, it seems preferable to predicate the design on the noise limits to be reached during 99% of the busy hour. Furthermore, since the behavior beyond this percentage is not so important, it is implied that the test signal would also be the maximum likely to be transmitted, and as such, would include effects due to overload distortion. The per-channel transmission level in the operative system also readily follows from the 1% curve of Fig. 8. A test signal corresponding to the long-term mean power of the distribution shown in Fig. 8, could give quite suitable information concerning nonlinear distortion, but the amplitude would be too small to produce overload effects, and, therefore, it would be of less value in a test procedure.

A simple relation exists which enables the expected signal-to-noise ratio S/N' (due to nonlinear distortion) in the operative system to be determined in terms of the white-noise test:[7]

$$\underset{\text{(weighted)}}{S/N'} = \frac{\gamma b}{X 1.5\ 10^3} N^{(1-\alpha)}$$

where γ is the signal-to-noise ratio measured with the white noise, b the channel spacing in cycles, and X and α are constants determined by the shape of the 1% curve of Fig. 8.

In this case

$X=0.32$
$\alpha=0.66$
$b=4{,}000$

Consequently,

$$S/N' = \gamma N^{0.34}\, 8.4 \tag{4}$$

Since the requirement prior to the application of pre-emphasis and de-emphasis networks is for a signal-to-unweighted noise ratio of 30.5 db, $S/N'=33.5$ db, and the result is

$$3.4 \log N = 24.2 - 10 \log \gamma \tag{5}$$

(second-order distortion predominant)

Equation 5 is, of course, a requirement related to the over-all fictitious circuit of reference. On the basis that the separate noise powers are additive, it is a straightforward matter to determine the requirement for any section or link. The expected long-term mean may be obtained for any specified number of channels (Fig. 9).

The white-noise test presents no special difficulty for equipments connected back-to-back, but there is great difficulty in interpreting results with the equipment on location. If a particularly long section is being tested, it is difficult to avoid the effects of fading, unless particularly favorable conditions are in effect, and there is no satisfactory way of isolating the nonlinear and thermal noise components. The only suitable technique is to test the shortest sections of the circuit during periods of the day when the signal strength is maximum so that the effects of thermal noise can be negligible. The noise powers of the separate links may then be added to estimate the total noise. This technique must also be treated with caution since it is possible that when sections are interconnected, voltage-adding products may become significant. Such a possibility may be determined by careful laboratory testing on the back-to-back equipment. By so doing, the law of addition can be obtained prior to installation and the over-all nonlinear noise for the operative system may be estimated.

Conclusions

The recommendations of the CCITT apply to virtually every phase of modern communications. Of particular interest to radio-relay engineering is the hypothetical reference circuit and the over-all signal-to-noise ratio. It is pointed out that the CCITT recommendations are primarily intended for international circuits and do not necessarily apply to all operational requirements. Distributions of multichannel signal power (associated with traffic) are given and it is shown that the nonlinear noise associated therewith, is dependent on the number of channels. The effect of the relative periods of maximum fading and maximum traffic and the proportion and periods of maximum international traffic, are also considered.

It is suggested that for the optional design of radio relays, the performance limits for international circuits might well be limited to the best 5% rather than the worst channels and that the number of channels on which the traffic volumes are based should exclude those intended for international circuits.

Thermal noise is determined principally by propagation conditions. During periods of worst fading, signal strength exhibits a Rayleigh distribution and there may be uncorrelated fading over a large part of the route. However, during this period, a significant improvement due to diversity operation can be achieved. Severe fading usually produces nonlinear noise due to multipath but at present this effect is not capable of simple prediction. The effects of the numbers of fading links and resultant thermal noise distributions are presented.

The absolute value of signal-to-noise ratio is examined on the basis of the increased operating time due to the loss of articulation in the presence of noise. Thus it is recommended that signal-to-weighted thermal or nonlinear noise should be greater than 40 db for 99% of any hour, excluding the contribution of multiplex equipment. From this and the noise distributions the requirements for a single link may be determined. The improvement due to the use of pre-emphasis and de-emphasis networks is also taken into account. These data lead to the generalized design requirement: the signal-to-thermal noise figure for a single unfaded link should be 64.5 db and the signal-to-nonlinear noise should be $[30.5+10 \log S]$ db where S is the number of homogeneous sections. These figures apply to the unweighted noise in the top channel prior to the use of base-band noise equalization and, within small limits, for systems of 10 to 50 links with second-order distortion predominant.

White-noise testing to determine nonlinear noise is a useful technique under controlled conditions. In the field, however, there may be much difficulty in separating thermal and nonlinear noise components. The preferable rms test-transmission level for an N channel system is $(-5.0+6.6 \log N)$ dbm0. The results obtained may be related easily to the operational system.

Appendix

The following notes are taken from the report of the U. S. Delegation to the Interim Meeting of the CCIR Study Group IX, Geneva, Switzerland, August 18–29, 1958. In particular, they refer to reference 43. They are included here because of their particular significance with respect to the contents of this paper and specifically wide-band radio-relay systems.

Noise aspects of radio-relay systems as related both to hypothetical reference circuits and to real circuits were considered by Subgroup IX-C.

A draft recommendation was prepared to replace Warsaw Recommendation 200. It specifies that the psophometrically weighted[*] noise power at a zero level point in any telephone channel on a 2,500-km (kilometer) hypothetical reference circuit for FDM systems should not exceed the following values (the assumption is made that the hypothetical reference circuits are loaded with uniform white noise at zero transmission level of $-15+10 \log_{10} N$ dbm for 240 channels or more, and $-1+4 \log_{10} N$ dbm for numbers of channels between 12 and 240, N being the total number of channels for which the radio relay system is to be designed):

1. 7,500 μμw mean value in any hour.
2. 7,500 μμw 1-minute-mean power for more than 20% of a month when the fading is severe.
3. 47,500 μμw 1-minute-mean power for more than 0.1% of a month when the fading is severe.

It was also decided that an additional item should be suggested to the Joint CCITT–CCIR Working Party on Circuit Noise for its guidance:

4. 1,000,000 μμw unweighted 5-millisecond-mean power for more than 0.01% of a month when the fading is severe.

It was recommended concerning noise in real circuits, that the psophometrically weighted noise power at a zero transmission level point in the telephone channels of a real FDM line-of-sight radio-relay system of length L km, where L is between 250 and 2,500 km, and for which composition does not differ appreciably from the hypothetical reference circuit, should not exceed $3L$ μμw mean value in any hour, or $3L$ μμw 1-minute-mean power for more than 20% of any month. Also as a planning objective (recognizing that the performance of real circuits may differ from the planning objective and is difficult to measure precisely), the afore-mentioned noise power should not exceed 47,500 μμw 1-minute-mean power for more than $(L/2{,}500)\times 0.1\%$ of any month.

Further study of permissible noise powers for systems of composition differing considerably from that of the hypothetical reference circuit is suggested.

A draft report was formulated on the noise tolerable during very short periods of time in radio-relay systems. Its annex is taken almost exactly from one of the documents submitted by the United States.

A statement was prepared giving the opinion of CCIR Study Group IX concerning noise objectives for radio-relay circuits carrying telegraph signals which was proposed by the Joint CCITT–CCIR Working Party on Circuit Noise. Specifically, it does not accept as an objective that 1,000,000 μμw unweighted-noise power at a zero transmission-level point with an integrating time of 5 millisecond must not be exceeded more than 0.001% of the time per month. It proposes instead, provisionally (as a noise objective for telegraphy), the limit mentioned previously of 1,000,000 μμw unweighted 5-millisecond-mean power for 0.1% of a month when the fading is severe.

---

[*] The level of uniform-spectrum noise power in a 3.1-kc band must be reduced by 2.5 db to obtain the psophometrically-weighted noise power.

References

1. The Organization and Functions of the CCIR, **E. W. Allen.** *Transactions*, Professional Group on Communications Systems, Institute of Radio Engineers, New York, N. Y., Nov. 1956, p. 26.

2. The Work of the International Telecommunication Union in the Field of Telephony, **W. E. Bloecker, G. A. Caldwell.** *AIEE Transactions*, vol. 76, pt. I, Nov. 1957, pp. 543–52.

3. XVII Plenary Assembly, CCIF, Geneva, Switzerland, vol. I-VI, Oct. 1954.

4. VIII Plenary Assembly, CCIR, Warsaw, Poland, vol. I-III, 1956.

5. *Ibid.*, vol. I.

6. General Telephone Statistics, 1955. International Telecommunication Union, 1956.

7. Design Factors for the Optimization of Multichannel Radio Systems, **Charles A. Parry** *AIEE Transactions*, vol. 76, pt. I, Nov. 1957, pp. 606–20.

8. XVII Plenary Assembly, CCIF, Geneva, Switzerland, vol. III, Oct. 1954.

9. *Ibid.*, vol. I.

10. Cross Modulation Requirements on Amplifiers Below Overload, **W. Bennett.** *Bell System Technical Journal*, New York, N. Y., Oct. 1940, p. 587.

11. Non-Linear Distortion in Transmission Systems, **R. A. Brockband, C. A. A. Wass.** *Journal*, Institution of Electrical Engineers, London, England, pt. III, Mar. 1945, p. 1.

12. Microwave Radio Links, **A. T. Starr, T. M. Walker.** *Proceedings*, Institution of Electrical Engineers, pt. III, Sept. 1952, p. 241.

13. Load Rating Theory for Multichannel Amplifiers, **B. D. Holbrook, J. T. Dixon.** *Bell System Technical Journal*, vol. 18, 1939.

14. The Effect of Non-Linear Distortion in Multichannel Amplifiers, **B. B. Jacobsen.** *Electrical Communications*, New York, N. Y., vol. 19, 1940, p. 29.

15. A Speech Volume Survey on Telephone Message Circuits, **V. Subrizi.** *Bell Laboratory Record*, New York, N. Y., Aug. 1953, p. 292.

16. XVII Plenary Assembly, CCIF, Geneva, Switzerland, vol. IV, Oct. 1954.

17. Design of FM Radio Relay Systems for Multichannel Telephone Service, **J. W. Halina.** *Transactions*, Professional Group on Communications Systems, Institute of Radio Engineers, Oct., 1956, p. 3.

18. RF Bandwidth of Frequency Division Multiplex Systems Using Frequency Modulation, **R. G. Medhurst.** *Proceedings*, Institute of Radio Engineers, Feb. 1956, p. 189.

19. Traffic Aspects of Communication Switching Systems, **J. A. Bader.** *Proceedings*, Eastern Joint Computer Conference Institute of Radio Engineers, Dec. 12, 1957.

20. Echo Distortion in the FM Transmission of Frequency Division Multiplex, **W. J. Albersheim, J. P. Schajer.** *Proceedings*, Institute of Radio Engineers, 1952, p. 316.

21. Intermodulation Distortion Due to Fading in Frequency Modulation Frequency Division Multiplex Radio Systems, **R. G. Medhurst, M. Hodgkinson.** *Proceedings*, Institution of Electrical Engineers, pt. C, Sept. 1957, p. 475.

22. Results of a Microwave Propagation Test on a 40-Mile Overland Path, **A. L. Durkee.** *Proceedings*, Institute of Radio Engineers, Feb. 1948, p. 197.

23. Notes on Propagation, **L. A. Byam, J. Z. Millar.** *Convention Record*, Institute of Radio Engineers, pt. 2, p. 68, 1953.

24. Statistical Data for Microwave Propagation, **P. Gudmandsen, B. F. Larsen.** *Acta Polytechnica*, Stockholm, Sweden, no. 13, Electrical Engineering Series, vol. 7, no. 7.

25. Realisation de Liaisons par Faisceaux Hertiziens de Qualite sur les Trajets Maritimes en Grece, **R. Cabessa.** *Onde Electrique*, Malakoff, Seine, France, Aug.-Sept. 1955, p. 714.

26. Some Features of VHF Tropospheric Propagation, **M. W. Gough.** *Proceedings*, Institution of Electrical Engineers, pt. B, 1955, p. 43.

27. A Statistical Study of Selective Fading of Super High Frequency Radio Signals, **R. L. Kaylor.** *Bell System Technical Journal*, Sept. 1953, p. 1187.

28. A Simplified Diversity Communications System for Beyond-the-Horizon Links, **J. Altman, W. Sichak.** *Transactions*, Professional Group on Communications Systems, Institute of Radio Engineers, Mar. 1956, p. 50.

29. An Evaluation of Radio Noise Meter Performance in Terms of Listening Experience, **C. M. Burrill.** *Proceedings*, Institute of Radio Engineers, Oct. 1942, p. 473.

30. The Influence of Amplitude Limiting and Frequency Selectivity upon the Performance of Radio Receivers in Noise, **J. C. R. Licklider, S. F. Goffard, W. J. Cunningham.** *Ibid.*, Oct. 1947, p. 1021.

31. Speech and Hearing in Communications (book), **H. Fletcher.** D. Van Nostrand Company, Inc., Princeton, N. J., 1953.

32. Study on the Effects of Noise on Speech Communications, **J. P. Egan.** *OSRD Report no. 2038*, Office of Scientific Research and Development, Washington, D. C., *PB 22907*, Nov. 25, 1943.

33. Masking by a Periodically Interrupted Noise, **I. Pollack.** *Journal*, Acoustical Society of America, New York, N. Y., Mar. 1955, p. 353.

34. Articulation Testing Methods, **J. C. Sheinberg, H. Fletcher.** *Ibid.*, supplement to Jan. 1930, p. 1.

35. A Theoretical Study of the Articulation and Intelligibility of a Telephone Circuit, **J. Collard.** *Electrical Communications*, Jan. 1930, p. 168.

36. Room Noise at Subscribers Telephone Locations, **D. F. Seacord.** *Journal*, Acoustical Society of America, July 1940, p. 183.

37. Systems Engineering of Tactical Multihop Tropospheric Scatter Circuits, **A. Mack.** *Technical Memorandum M1728*, Signal Corps Project no. *807B*, Signal Corps Engineering Laboratories, Redbank, N. J., May 1956.

38. The Equalization of Baseband Noise in Multichannel FM Radio Systems, **C. A. Parry.** *Proceedings*, Institute of Radio Engineers, Nov. 1957, p. 1527.

39. Interchannel Interference in FM and PM Systems under Noise Loading Conditions, **W. R. Bennett, H. E. Curtis, S. O. Rice.** *Bell System Technical Journal*, May 1955, p. 601.

40. An Extended Analysis of Echo Distortion in the FM Transmission of Frequency Division Multiplex, **R. G. Medhurst, G. F. Small.** *Proceedings*, Institution of Electrical Engineers, pt. B, Mar. 1956, p. 190.

41. Noise Considerations on Toll Telephone Microwave Radio Systems, **T. A. Combellick, M. E. Ferguson.** *AIEE Transactions*, vol. 76, pt. I, Mar. 1957, pp. 67–70.

42. Some Design Considerations for Links Carrying Multichannel Telephony, **S. Fedida, D. S. Palmer.** *Marconi Review*, St. Albans, Hertfordshire, England, vol. XVIII, 1955, p. 132, and vol. XIX, 1956, p. 1.

43. *Document no. 20, TD Serial no. 890.* Study Group IX, CCIR, Geneva, Switzerland, Oct. 1958.

44. Distortion in Frequency Division Multiplex Systems Due to an Interfering Carrier, **R. G. Medhurst, E. M. Hicks, W. G. Grossett.** *Proceedings*, Institution of Electrical Engineers, pt. B, no. 21, May 1958.

Marginal Utility and a Criterion of Performance for Communication and Radar Systems

L. S. SCHWARTZ
MEMBER AIEE

SINCE communication or radar systems, like all systems involving human activity, are subject to the laws of profit and loss, their performance may be evaluated on a scale of economic utility values. A utility evaluation may be conducted with the aid of D. Bernoulli's Law of Diminishing Marginal Utility.[1]

The marginal utility law orginated in reflections by Bernoulli to the effect that the principle of maximizing the estimated economic gain could lead to unfavorable results in some cases. For example, if the possible loss for X is not small in relation to his fortune, he should not choose an even-money bet, although the principle of maximizing the estimated gain would recommend this. The reason, as Carnap[2] states, is that "the weight of a large loss should be more than proportional because X would suffer from a large loss disproportionately. If X has a fortune of 10,000, then he would suffer from a loss of 8,000 not only eight times as much as from a loss of 1,000, but much more because it would mean his near ruin. If X were to lose by ten successive accidents 1,000 each time, then every loss would hurt him more than the preceding ones, and the last would be the worst. Inversely, if X, with an initial fortune of zero, were to make ten or any other number of successive gains of 1,000 each, the satisfaction derived from subsequent gains would be smaller and smaller." Following the terminology of economists, the satisfaction conveyed by money or goods is measured by a number called its utility. Therefore, X's decisions must be guided by the principle of maximizing the utility of his gains rather than the gains themselves. Since, however, he cannot in general anticipate gains and utilities with certainty, he must apply the maximizing principle to the estimate of utility rather than to the unknown utility itself.

These ideas led Bernoulli to formulate the change in utility du with change in fortune df as

$$du = k\frac{df}{f}$$

where k is a constant which is characteristic of the person X at the time in question.

In general, since the fortune change cannot be known with certainty, i.e., it will be one of several possibilities with associated probabilities, the estimated utility of a fortune change only can be stated. Thus, suppose that f_0 is the initial fortune and that f can take on one of two values, f_1 or f_2, with probability q_1 or q_2, respectively. Then

---

Paper 58-1264, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted January 3, 1958; made available for printing January 7, 1959.

L. S. Schwartz is with New York University, New York, N. Y.

$$u^{*1}=k(q_1 \log f_1+q_2 \log f_2-\log f_0) \tag{1}$$

where u^{*1} is the estimated utility value of the fortune change. Application of this idea to communication systems may take the form of assuming that an observer makes decisions with the aid of received information and that these decisions are rewarded or penalized. If the decisions are correct, his initial fortune is increased; if they are incorrect, it is decreased. The observer determines for himself the amount he gains or loses each time he makes a decision. If the decision is correct, the fortune following his decision is

$$f_1=(1+L)f_0 \tag{2}$$

where L is the percentage of his initial fortune invested in the decision. If the decision is incorrect, the fortune following the decision is

$$f_2=(1-L)f_0 \tag{3}$$

In these expressions gains and losses are given equal weight, but this is not necessary.

Assuming that the probability of correct decision is q and that of incorrect decision p and letting $q_1=q$, and $q_2=p$ in equation 1, the estimated utility of the fortune change becomes

$$\begin{aligned} u^{*1}&=k[q \log (1+L)f_0+p \log (1-L)f_0- \\ &\qquad \log f_0] \\ &=k[q \log (1+L)+p \log (1-L)] \end{aligned} \tag{4}$$

The value of L, which maximizes u^{*1} is

$$L=q-p,$$

so that

$$u^{*}=k[1+p \log p+q \log q] \tag{5}$$

which, except for the constant factor k, is the expression for the capacity of a symmetric binary communication channel of error probability $p(q=1-p)$.

Kelly[3] obtained the expression for the capacity of a symmetric binary communication channel by postulating a situation in which a gambler is receiving tips over a noisy communication channel. If the input symbols to the communication channel represent the outcomes of a chance event on which bets are available at odds consistent with their probabilities, the gambler can use the knowledge derived from the received symbols to cause his money to grow exponentially. The maximum exponential rate of growth of the gambler's capital is equal to the rate of transmission of information over the channel. Kelly's approach is suitable for the situation in which there is no limitation on the number of bets which can be placed and in which there is no requirement to maximize the reliability of betting. These limitations do not inhere in the utility value approach whose usefulness is demonstrated in the two following cases.

Case 1. Information Rate of Communication Systems

If repetition-integration is employed,

$$u^{*}=\frac{k}{(r+1)}[1+p \log p+q \log q] \tag{6}$$

with p a monotonically decreasing function of the integrations r. As $r \to \infty$, $q \to 1$, and $p \to 0$, and in the limit the entire initial fortune may be safely invested in each decision. However, the number of possible decisions varies inversely with the repetitions, and for an infinite number of integrations there are no decisions. Although the existence of a value of r which maximizes u^* might be expected, for Gaussian noise such a value does not exist because the quantity in the square brackets is a monotonically increasing function of r with maximum value of unity and with downward concavity. Hence u^* can only decrease with r. It appears, therefore, that instead of increasing the reliability of a communication channel by repetition-integration, one should simply invest that fraction of his initial fortune which corresponds with the reliability of the channel. In communication language this means that if time is unlimited, one should operate with maximum information rate. In this way the estimated utility is a maximum, a result consistent with the concept that the information rate is bounded by the capacity of the channel.

Case 2. Reliability of Radar Systems

To see the implication of the utility value concept to radar, it is useful to distinguish between certain characteristics of radar and communication systems. Communication is a co-operative enterprise between the transmitter and the receiver. Its chance nature derives from the presence of interference in the channel, originating from either natural or man-made causes. Radar, on the other hand, is essentially a measurement system in which observations are taken in an environment of noise. Because of the co-operative nature of communications, the receiver is directly penalized only if he makes a wrong decision concerning the transmitted signal. If he withholds decision until he is more confident, he makes fewer decisions and thereby reduces the possible gain which he might make by making more decisions. Only in this sense does he lose by waiting. In radar, on the other hand, the observer may lose equally by waiting or by deciding. If the target is an enemy bomber, it matters little whether the decision is that there is no target, or whether decision is withheld. Aside from losses which may be inflicted on enemy planes, the radar observer must be primarily concerned with safeguarding his initial fortune f_0. It may be lost by making a wrong decision or by failing to take one, either way.

Hence in the radar case the fortune following decision is either f_0 or βf_0, $0 \leq \beta \leq 1$, where β is a factor relating to the cost of a wrong decision, decreasing with increasing cost.

Thus,

$$f_1=f_0 \tag{7}$$

and

$$f_2=\beta f_0$$

so that the estimated utility is

$$\begin{aligned} u^{*}&=k[q_1 \log f_1+q_2 \log f_2-\log f_0] \\ &=[q_1 \log f_0+q_2 \log \beta f_0-\log f_0] \\ &=k[q_2 \log \beta] \\ &=k[p \log \beta] \end{aligned} \tag{8}$$

As $\beta \leq 1$, $u^* \leq 0$. For β small, u^* will be a correspondingly large negative number, indicating excessive loss. This can be avoided only by integration which reduces p, thereby increasing system reliability. In fact, as r, the number of integrations, becomes infinite, $p \to 0$. For radar $u^*_{\max}=0$. However, in practice u^* must be less than $u^*_{\max}$ because the number of integrations must remain finite. In fact, the number of integrations is limited by the data rate requirement, that is, by the time available for decision.

Thus, a radar system is not concerned with information rate, since its estimated utility is independent of information rate, increasing solely with system reliability.

Conclusion

The law of diminishing marginal utility can be applied not only to phenomena in the economic field, but also to those in the fields of communication and radar, a consequence of the fact that communications and radar involve decisions possessing an economic value measure. In the case of communication with no limitation on the number of decisions, i.e., on message duration, the criterion of performance is information rate because maximizing the information rate maximizes the estimated utility. In radar maximizing the reliability maximizes the estimated utility, consistent with data-rate requirements, and thus information rate is not a criterion of radar performance.

References

1. Specimen Theoriae Novae de Mensurae Sortis, D. Bernoulli. *Commentarii*, Academie Science Imperialis Petropolitanae (for 1730 and 1731), vol. 5, 1738, pp. 175–92.

2. Logical Foundations of Probability (book), R. Carnap. University of Chicago Press, Chicago, Ill., 1950.

3. A New Interpretation of Information Rate, J. L. Kelly, Jr. *Bell System Technical Journal*, New York, N. Y., vol. 35, July 1956, pp. 917–26.

The 83B1 Teletypewriter Selective-Calling System

C. W. SMITH
ASSOCIATE MEMBER AIEE

C. J. VOTAW
NONMEMBER AIEE

A. L. WHITMAN
MEMBER AIEE

During the last decade and a half in particular, the use of teletypewriters on leased lines has grown very rapidly. The Bell System now has more than 50,000 such teletypewriter stations in service. In order to obtain the most communication for the least over-all cost, it is common practice, where the traffic load on the line will permit, to operate several stations on the same line. This results in a significant reduction in line cost compared with a line facility for each pair of stations. However, such multistation lines involve operating complications which somewhat offset the lower cost.

With several stations connected to the same party line without selecting equipment, all stations will receive all messages, and any sender must wait until the line is idle before he tries to send. An attempt at simultaneous sending may result in hopelessly garbled messages. Furthermore, even if a sender does wait until the line is idle, some other sender may also be waiting and both may attempt to send at about the same time.

In certain cases, circuit contention may not be a serious problem and a simple party line without station selection may serve in a satisfactory and economical manner. In many cases, however, such a simple party line would not be satisfactory. For example:

1. It may be desired to direct a message to one station, or a selected group of stations only. This may be done in order to obtain a certain degree of privacy, or more commonly, to obviate the necessity of checking all the traffic received at each station to select the messages intended for that station.

2. It may be important to eliminate contention for the opportunity to use the circuit. This contention can result in a certain amount of garbled copy which would require repetition, and additional lost time in coming to an agreement as to who is to use the circuit.

These points are especially important where the traffic load is high or where automatic tape sending is desired.

A selective-calling arrangement will direct messages to the particular stations for which they are intended. A number of types of selective-calling arrangements for use on teletypewriter circuits have been widely used. Various types of anticontention arrangements have been devised to prevent the confusion resulting from simultaneous attempts by two or more stations to send. The most successful of these is an automatic-transmitter-start arrangement. With this feature, the transmitter or keyboard at each station is under the control of one station on the line, the control station. The control station gives permission to each of the stations to send, in a predetermined order, by means of selective signals.

The *81*-type teletypewriter switching systems[1] provide these features for multistation full-duplex lines in a multiline system. These systems provide very complete means for automatic switching of messages together with alarms and elaborate means for supervising systems operation.

The *83B1* teletypewriter selective-calling system is intended to fill the need for a less elaborate and less costly system where station selection and the elimination of circuit contention are important. It is primarily adapted to systems consisting of a single-party line, but it may be used in multiline manual-tape-relay systems and also for automatic interline switching for a limited number of lines.

The *83B1* system combines the use of a particular type of station-selecting arrangement with an automatic transmitter-start feature. The most common station equipment includes an automatic sending and receiving (ASR) teletypewriter set, which includes, in addition to a teletypewriter for receiving and keyboard sending, a keyboard perforator for punching tape and a transmitter-distributor for sending from punched tape. Fig. 1 shows a *19ASR* set as installed at an *83B1* control station. The *19ASR* set was offered as an interim arrangement until a new ASR set in the *28* line (manufactured by the Teletype Corporation) became available, when some simplification of the *83B1* arrangements was possible.

The system also may be used with other teletypewriter equipment with some reduction in the features provided, or with the use of separate units to provide required features.

Station selection is obtained from the *28*-type stuntbox,[2] either in a printer or in a device called a sequence selector, which is a printer with the printing and paper-handling features removed, leaving only the selection features. In a printer the stuntbox provides means for operating various functions, or stunts, such as

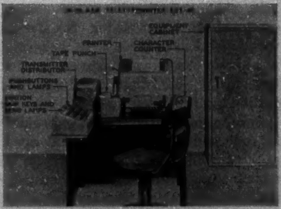


Fig. 1. Control station with push-button calling

Paper 58-443, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 1, 1957; made available for printing November 13, 1958.

C. W. Smith, retired, was formerly with the American Telephone and Telegraph Company, and C. J. Votaw and A. L. Whitman are with the Bell Telephone Laboratories, Inc., New York, N. Y.

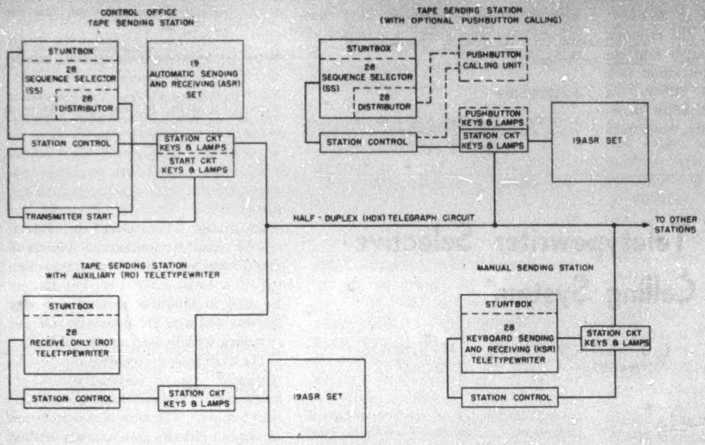

CONTROL OFFICE
TAPE SENDING STATION
STUNTBOX
28 SEQUENCE SELECTOR (SS)
28 DISTRIBUTOR
19 AUTOMATIC SENDING AND RECEIVING (ASR) SET
STATION CONTROL
STATION CKT KEYS & LAMPS
START CKT KEYS & LAMPS
TRANSMITTER START
TAPE SENDING STATION
(WITH OPTIONAL PUSHBUTTON CALLING)
STUNTBOX
28 SEQUENCE SELECTOR (SS)
28 DISTRIBUTOR
PUSHBUTTON CALLING UNIT
PUSHBUTTON KEYS & LAMPS
STATION CKT KEYS & LAMPS
STATION CONTROL
19ASR SET
HALF - DUPLEX (HDX) TELEGRAPH CIRCUIT
TO OTHER STATIONS
TAPE SENDING STATION
WITH AUXILIARY (RO) TELETYPEWRITER
STUNTBOX
28 RECEIVE ONLY (RO) TELETYPEWRITER
STATION CONTROL
STATION CKT KEYS & LAMPS
19ASR SET
MANUAL SENDING STATION
STUNTBOX
28 KEYBOARD SENDING AND RECEIVING (KSR) TELETYPEWRITER
STATION CKT KEYS & LAMPS
STATION CONTROL

Fig. 2. Types of 83B1 system stations

figures shift (FIGS), letters shift (LTRS), carriage return (CR), and line feed (LF). In addition, the stuntbox can be arranged to:

1. Shift the typing unit into a nonprint condition upon the receipt of a suitable character combination, such as FIGS H LTRS.
2. Shift the typing unit back into the print condition upon the receipt of a selected character or combination of characters.
3. Open or close contacts upon the receipt of selected characters or character combinations. Such contact operation may be momentary, or the contacts may lock in the operated position and be released by a subsequent character combination.
4. Lock the stuntbox upon the receipt of an end-of-address or lock-out code, such as CR LF LTRS, so that no further selections may take place except for the shift to the nonprint condition and the release of all locked-in selections.

By these features, station selection is obtained to start transmitters individually and to call in one or more stations to receive a message.

In addition to the stuntbox, a station control circuit is provided at each station which functions in response to the operation of contacts in the stuntbox to control the starting of transmitters and the cutting on of receivers.

One transmitter-start circuit is provided for each multistation line. It is located at the control station. Its function is to start the transmitters at the various stations, one at a time, provided they have traffic to send. At manual keyboard stations without automatic tape transmitters, the attendant is alerted, provided she has indicated that she has traffic to send, and allotted a 10-second interval in which to start sending from the keyboard.

Features Provided

Figs. 2, 3, and 4 show typical *83B1* station arrangements in block diagram form.

The system provides features described in the following paragraphs.

Adaptation to Various Types of Teletypewriter Apparatus

The following teletypewriter apparatus may be used:

19ASR teletypewriter set

28ASR set

28 receiving only (RO) printer

28 keyboard sending and receiving (KSR) teletypewriter

14 typing reperforator and *14* transmitter-distributor

14 reperforator-transmitter (for interline switching)

Stations which are arranged to send automatically teletypewriter characters for selecting purposes without first having them punched in tape, are equipped with a *28*-type distributor. This distributor can be coded to send various letters by means of external circuits, but is not arranged to send from punched tape.

Tape Preparation and Format

At a station equipped for tape sending, the message is prepared in the form of punched tape. At the beginning of the tape, one or more call-directing codes

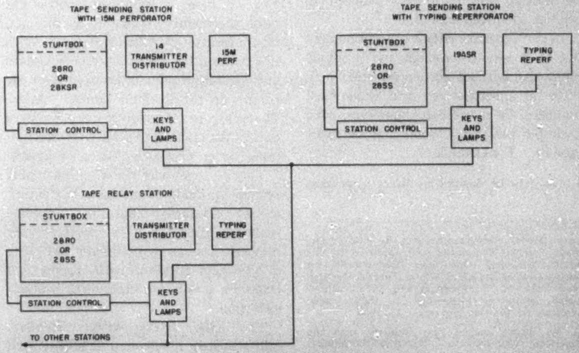

TAPE SENDING STATION
WITH 15M PERFORATOR
STUNTBOX
28RO OR 28KSR
14 TRANSMITTER DISTRIBUTOR
15M PERF
STATION CONTROL
KEYS AND LAMPS
TAPE SENDING STATION
WITH TYPING REPERFORATOR
STUNTBOX
28RO OR 28SS
19ASR
TYPING REPERF
STATION CONTROL
KEYS AND LAMPS
TAPE RELAY STATION
STUNTBOX
28RO OR 28SS
TRANSMITTER DISTRIBUTOR
TYPING REPERF
STATION CONTROL
KEYS AND LAMPS
TO OTHER STATIONS

Fig. 3. Types of 83B1 system stations

(CDC's) are punched. The CDC's are 2-letter codes. Each station is assigned a CDC to which it alone will respond. The CDC's of several stations may be inserted in the tape ahead of the message if it is intended for more than one station. To speed operation with multiaddress messages, predetermined group codes may be assigned for groups frequently called, as well as a broadcast code to include all stations on the line. After each CDC, an LTRS character is inserted. After all the address CDC's have been punched, the end-of-address, or lockout code, CR LF LTRS, is punched. This is followed by the text of the message and finally, the disconnect (end-of-message) code, FIGS H LTRS. A typical message format is shown in Fig. 5.

When the message tape is completed, it is torn off and inserted in the gate of the transmitter with the tight tape arm lowered. This is done whether the line is busy or not. The BID lamp lights. If desired, the station may be operated with continuous tape. In this case the end-of-message code of the first message must be followed immediately by the second message or by about 100 LTRS characters punched in the tape, so that there will be enough tape available to permit all of the first message to go through the transmitter head.

Transmitter Start

At the control station, the transmitter-start circuit functions to start the transmitters in order, provided they have traffic to send. This is done by sending transmitter-start codes (TSC) to each station. Each TSC is a 2-letter code like the CDC, but is distinguished from a CDC by the choice of the first letters.

The TSC's are stored in the transmitter-start equipment and sent one at a time in a predetermined order. When the station receives a TSC in its stuntbox and has no traffic to send, it transmits back to the control station the letter *V* as a no-traffic response.

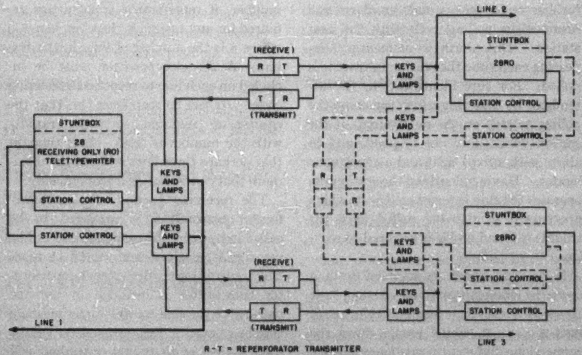
The receipt of this *V* at the control station is a signal to poll the next station. If a station has traffic to send, there is no *V* response. The transmitter starts automatically and sends the message.

After completion of a polling cycle the entire process is repeated. If desired, certain transmitters with anticipated heavier traffic may be polled more than once in a cycle.

Supervision of the polling operation is given to the attendant at the control station by providing a station lamp and a skip key for each station. A lighted lamp indicates that a station is being polled, or, if the lamp remains lighted, that the station is sending. The skip key allows the attendant to turn off the TSC for that particular station so that it is not polled while "on skip." This may be used when a station is taken out of service for maintenance, or, in connection with the polling patterns described hereafter, to control the flow of traffic from different stations.

Polling Patterns

The particular codes used for polling the stations should, in general, be of no interest to the customer and therefore may be selected for convenience in equipment arrangement. It has been thought that a number of specific arrangements would be particularly useful. The equipment may be modified by simple plug and cord changes for any of five arrangements designated polling patterns *A*, *B*, *C*, *D*, and *E* as follows:



R-T = REPERFORATOR TRANSMITTER

Fig. 4. Automatic relay station

```
(10 LTRS) AD LTRS AE LTRS BE LTRS BR LTRS
CR LF LTRS (MESSAGE TEXT) FIGS. H LTRS
(ADDITIONAL LTRS AS NEEDED FOR TEARING TAPE
OR FOR CONTINUOUS - TAPE OPERATION)
```

Fig. 5. Typical message format for multiple-address message

Polling Pattern *A*. This pattern can be used for a maximum of ten stations. Each station has one skip key and each station is polled four times for each revolution of the selector switch. More efficient use of the switch is the result.

Polling Pattern *B*. This is also intended for 10-station lines. Each station is polled four times per cycle and each station has two skip keys. Thus a station may be skipped entirely or twice out of four possible pollings.

Polling Pattern *C*. This pattern is for a maximum of 20 stations. Each station is polled twice in a revolution and each has one skip key.

Polling Pattern *D*. This pattern is also for a maximum of 20 stations. Each station is polled twice per revolution and has two skip keys.

Polling Pattern *E*. This pattern can be used for a maximum of 36 stations. Each station is polled only once per revolution, and has only one skip key.

If none of these specific patterns are satisfactory the selector can be wired to order for any desired pattern.

Priority Traffic

Urgent messages may be given priority so that they will be sent ahead of others. Each station has a priority button which, when pushed by the attendant, gives priority to the next message. Priority is obtained by a feature of the station-control circuit such that a transmitter with tape ready to send does not start until it has been polled twice unless its priority button has been pushed. Thus the polling cycle consists of two complete rounds of polling with the same TSC's, the first a search for priority traffic and the second a search for regular traffic. After every message sent, this pattern is repeated starting with the TSC for the next station in line. If no priority buttons have been pushed, no station starts on the first round of polling.

For types of operation where priority is not needed and the extra polling time is not justified, the priority option can be removed by a circuit strap at each station.

The sending of a priority message cancels the priority status of the station sending, so that if priority is wanted on a following message, the button must be pushed again. However, once a priority button has been pushed, the status is held until the message is sent from that station; it is not cancelled by the use of priority by another station which happens to get the first chance to start in the polling sequence.

Sending the Message

When a station having traffic to send is polled (the first or second time, depending on the priority status) the transmitter starts and sends first the CDC or CDC's which direct the message to the proper stations. At each called station, but at no others, the printer is unblinded (if in a *19ASR* set) or shifted into the print condition (if a *28* printer) on the receipt of the CDC. The sending of the CDC stops the transmitter at the sending station, which waits for a *V* answerback from the called station. This answerback is also the letter *V*, but its function differs from the *V* sent as a no-traffic response. This letter *V* now indicates that the printer at the called station is ready to receive a message. The arrival of the *V* at the sending station restarts the transmitter for the sending of the rest of the message.

After all of the CDC's have been sent, the end-of-address or lock-out code, CR LF LTRS, is sent. This code has two functions. It returns the carriage to the left and feeds out the paper so that the message text starts at the left side. In addition, it operates a mechanism in all of the stuntboxes on the line so that subsequent letter combinations in the text which are the same as CDC's, do not select any unwanted stations. Then follows the text of the message.

At the end of the text the disconnect signal, FIGS H LTRS, is sent from the tape. This disconnects all connected stations and causes the transmitter-start circuit to resume polling after a delay of about 300 milliseconds.

Keyboard Sending and Receiving Stations

At keyboard sending stations there is no automatic tape transmitter. Such stations are polled by sending TSC codes, however, like tape-sending stations. The station attendant indicates that there is traffic to send by pushing the KS (keyboard send) button on the control panel, which lights the BID lamp. This is equivalent to placing tape in the transmitter at a tape-sending station. When the station is polled, if the attendant has not made a bid, a *V* no-traffic response is sent back to the central office. If there is traffic to send, an LTRS character is sent back which simulates the start of sending and stops polling by the transmitter-start circuit. An audible alarm is sounded at the station, the station SEND lamp is lighted and the BID lamp goes out. These two signals tell the station attendant that she has 10 seconds to start sending. If there is no sending in 10 seconds, the transmitter-start circuit automatically will send FIGS H LTRS U and start a new polling cycle. The sending station will lose the opportunity to send, and must wait until the station is polled again. The same thing will happen if the attendant stops sending during a message for a period of 10 seconds.

Circuit Assurance

The *V* answerback response to a CDC, and the *V* no-traffic response to a TSC, constitute an effective circuit-assurance feature. Each time a TSC is sent to poll a station for traffic, either a *V* is returned indicating that the TSC has been received and there is no traffic to send, or the transmitter starts and sends the message that is waiting. In either case there is reasonable assurance that the line facility and the equipment are in operating condition. Likewise with the *V* answerback to a CDC, reception of the answerback is reasonable assurance of an operable circuit.

Timeout and Alarms

The system is provided with several features to prevent its automatic operation from being stopped by trouble affecting only part of it, such as one or more individual stations. These features cause an alarm to indicate the location of the trouble while operation of the rest of the system proceeds normally. Protection is given against failure of a response to a TSC, failure of a response to a CDC, and interruption of transmission from any cause.

If no response is received from a TSC, the transmitter-start circuit times out for 2 seconds, then sounds an alarm and resumes the polling cycle with the next station. This alarm is momentary, repeating each time the station in trouble is polled. For sure identification, the attendant, by throwing a key, can cause the polling to stop on the next round, at the station which does not respond, and to alarm with a fixed light and a continuous buzzer. Having identified the station she operates the skip key assigned to it, which prevents it from being polled until the trouble is found and it is restored to service.

If no answerback is received from a CDC, the transmitter-start circuit waits 10 seconds and then automatically sends FIGS H LTRS U, which breaks down the connections and conditions the circuit for a resumption of polling. At the sending station the red alarm lamp (AL) lights and the station transmitter cannot start again until the alarm is released. This warns the operator that she must examine the tape in the transmitter, correct the trouble if possible, and reset the tape to the beginning of the message before releasing the transmitter to start again.

If for any reason at all, the line remains idle and marking for a 10-second period, the same time out occurs and circuit operation, excluding the offending station, continues normally.

Manual Tape Relay

Any station on an *83B1* line may be used as a manual tape-relay station for relaying messages to and from another line. The *83B1* system then becomes a multiline system.

A typing reperforator, with or without an associated teletypewriter, is arranged to be called in by a suitably assigned CDC. The message for relaying to another line is received in punched tape. The tape is torn off and carried manually to the transmitter of the line for which it is intended and sent.

CDC's for stations on another line should be nonvalid for the line on which they were originally sent. For this reason, different sets of CDC's should be used on lines between which there is to be relaying of messages.

Automatic Interline Switching

Any station on an *83B1* line may also be used as an automatic interline switching point for the interchange of traffic with another similar line or with more than one line. The relaying of messages is accomplished by means of reperforator transmitters. Each line requires one reperforator transmitter for each line in the system other than itself. The total number of reperforator transmitters required for any system is therefore $n(n-1)$ where n is the number of lines in the system. A *28* teletypewriter must be included on each line to respond to selecting codes. It will be seen from this that the equipment required increases rapidly with the number of lines. It is thought that perhaps three lines is about the maximum that can be handled economically.

The receiving portion of each reperforator transmitter is arranged to be called in by a suitably assigned CDC. The message is received and stored in tape. The transmitter portion of each reperforator transmitter is included in the TSC pattern of the line to which the message is to be relayed. This transmitter will be started then in its proper order and will send the stored message.

Care must be taken in assigning CDC's where interline switching is provided. Codes assigned to stations on other lines should be nonvalid for the line on which they originate.

KEYS AND LAMPS

Lamp and audible signals are provided that indicate, both at the control point and at outlying stations, those conditions which need to be known to permit proper operation and adequate supervision of the traffic. Key switches are also provided to control circuit operations. Following is a list of the keys and lamps normally provided, with a brief statement of the function of each. The term key is used for a locking twist key, the term button for a nonlocking push button.

At the Control Station Only

The Station Lamp indicates which station is being polled or is sending.

The Station Skip Key controls the sending of individual TSC's.

The *V* Test and Stop Key control operation of the transmitter-start circuit.

The Alarm Transmitter Start Lamp (ALT) lights to indicate failure to receive a response from a polled station.

At All Stations (Including Control Station)

The BID Lamp lights to show that a bid to send has been made by inserting tape at an automatic sending station, or by pushing the KS (keyboard send) button at a keyboard station.

The KS Button conditions a station for keyboard sending when polled.

The SEND Lamp lights while the transmitter is running. At keyboard stations, it shows attendant when to send.

The MS (Manual Send) Button gives manual control of transmitter or of keyboard, without a wait for polling.

At All Stations

The PR (Priority) Lamp lights when the PR button is pushed to show a bid for priority.

The PR Button enters a bid for priority sending.

The ACO (Alarm Cutoff) Lamp when lighted indicates that the buzzer has been silenced by the AG (alarm guard) key.

The AG Key controls the alarm buzzer.

The AL Lamp (red) may be found at each station. When lighted, it indicates that the transmitter is locked out and will not start again until the alarm is released.

The AR (Alarm Release) Button puts out the alarm lamp and unlocks transmitter.

The MR (Monitor) Key permits the local typing unit to monitor or to print home copy automatically (with the *19ASR* set only).

PUSH-BUTTON SELECTION

Optional equipment is provided to permit station CDC's to be sent by operating push buttons instead of punching tape. If a message is to go to a number of stations, all of the required CDC's may be set up in the push-button circuit before the tape is inserted in the transmitter. They will then be sent, without further attention from the operator, ahead of the material in the tape, when the station next receives a proper poll. The CDC's, the same as those which would otherwise be punched in the tape, are generated by relays and a stepping selector and sent by an auxiliary cam-operated distributor driven by the motor of the typing unit or sequence selector.

Push-button selection may be provided at any station on an *83B1* system. Its use will result in a saving in time and a reduction in operator errors which might occur with manual punching of directing codes in the tape. It is especially valuable when manual tape-relay operation is desired, particularly for the relay point, where tapes without addresses may be received from other lines.

HALF- AND FULL-DUPLEX OPERATIONS

The *83B1* system is for half-duplex service only. A system for full-duplex operation is currently being designed.

SPEEDS

The system is capable of operating at 60, 75, or 100 words per minute, provided that the associated teletypewriter equipment is capable of operating at the same speed.

SYSTEM CAPACITY

On any one line the number of stations permissible will generally be limited within the capabilities of the system by the traffic generated and the speed of service required. The number of letter combinations for TSC's and CDC's which the system is designed to use sets an upper limit on the number of stations.

The circuits of the system have been designed to operate with 2-letter TSC's and CDC's. Also, in the interest of economy and simplicity in control circuits, certain limitations have been imposed on the choice of letter combinations.

Fig. 6. Outlying station with auxiliary receiving-only 28 printer

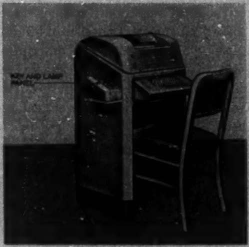


Fig. 7. Outlying keyboard sending and receiving station

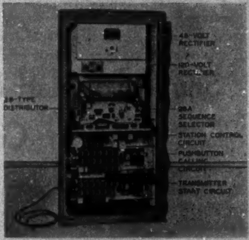


Fig. 8. Control station equipment cabinet

In general the letters *T*, in particular, *O*, and *M* are least desirable for selective-calling codes because, since each of these letters can be produced by a single open on the circuit, they could be produced falsely by hits on the line. The letter *H* is undesirable for a TSC or CDC since it is used in the disconnect code. The letter *V* is used for no-traffic response and answerback. The letter *Y* is not used in TSC's. This leaves 20 letters for TSC's. In the circuit design, the first letter of any TSC is restricted to either *Q* or *X*. The use of *Q* and *X* with each of the 20 letters as a second letter, including such combinations as *QQ*, etc., would result in 40 TSC's.

However, a restriction is imposed on the

TSC assignment by the system operation with CDC's. On any one *83B1* system, the first letter of all CDC's must be limited to one of two letters. It is the first letter of the CDC which stops the transmitter so that it waits for an answer back from a receiving station being called. To make it simpler for the system to distinguish between TSC's and CDC's, the two letters used for the first letter of the CDC's on a particular line may not also be used as the second letters of TSC's. As an illustration, if *A* and *B* are the first CDC letters for a line, *QA*, *QB*, *XA*, and *XB* cannot be used for TSC's on that line. Thus the total number of TSC's is reduced to 36, so that the total number of sending stations on a line is 36. This description applies to the standard wiring of the transmitter-start circuit and push-button circuit. The restriction is a practical, not a theoretical one. With changes in the strapping of the stepping selector, and with the use of *Y* or possibly the less desirable *T*, *O*, or *M*, the line capacity could be increased. However, it is expected that the demand for as many as 36 stations on such a system will be very small.

With CDC's, the letter *Y* need not be eliminated. So, with 21 second letters and two first letters, 42 different CDC's are available for a line where codes are sent from tape. Push-button codes are limited by the equipment to 20 second letters. *Y* is not used.

Station Equipment Arrangements

At *83B1* stations, in addition to the teletypewriters required, there are units of control equipment consisting of relays, stepping switches, electron tubes and associated equipment, together with power supplies, all mounted in apparatus cabinets. These apparatus cabinets can be located either in the same room with the teletypewriters or in an adjacent room as desired. Also fastened on the teletypewriter table, or *28*-type cabinet, are small metal cabinets in which are mounted the various signal lamps, control keys, and push buttons used in the operation of the system. Fig. 1 shows a typical control station equipped for a 20-station line and provided with push-button calling. The equipment includes the *19ASR* teletypewriter set, a 4-foot cabinet to house the sequence selector and the relay equipment, and the key and lamp cabinets shown mounted on the table.

Fig. 6 shows a typical outlying station with an auxiliary *28* RO teletypewriter. With this machine to print incoming copy, the *15* printer can be used to make copy while tape is being punched. At this station the control equipment is housed in the teletypewriter cabinet.

Fig. 7 shows a keyboard sending station with a *28*-KSR teletypewriter. Here also the control panel is mounted in the cabinet.

Fig. 8 shows the assembly of equipment in the cabinet for the control-station arrangement of Fig. 1.

The *83B1* system in expected to fill a growing need for improved teletypewriter service in cases where station selection on party lines, with the elimination of circuit contention, is important. The system has particular appeal because of its flexibility and its consequent capability of being adapted to many diverse uses.

References

1. A FULL AUTOMATIC PRIVATE-LINE TELETYPEWRITER SWITCHING SYSTEM, W. M. Bacon, G. A. Locke. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 473–80.

2. SIMPLIFIED PRINTING TELEGRAPH SWITCHING AND INTEGRATED DATA PROCESSING, J. B. Booth, R. H. Klich. *Ibid.*, vol. 75, pt. I, Mar. 1956, pp. 11–14.

3. AUTOMATIC DISPATCH SYSTEM FOR HALF-DUPLEX TELETYPEWRITER LINES, J. L. Maxwell, H. K. Farrar. *Ibid.*, Sept., pp. 441–45.

Vol.
Pt. 1
1959

Shunt-Coupled Magnetic-Amplifier Circuits

R. M. HUBBARD
ASSOCIATE MEMBER AIEE

A RECENT STUDY of magnetic-amplifier theory based on block-diagram techniques[1] suggested to this author that one potentially useful basic technique for extracting the output information from the simple self-saturating magnetic amplifier had been seriously neglected in the past. Although it is a well-recognized fact that the output information in a self-saturating magnetic amplifier is contained in the half-cycle average voltage across the gate winding, independent of power-supply variations, most of the conventional magnetic-amplifier circuits are so arranged as to make the output voltage proportional to the difference between the average gate voltage and the average supply voltage. (Subsequent to the preparation of this paper, Rosenstein published a description of a transactor[2] which bears a family resemblance to the shunt-coupled circuit described in this paper.) The afore-mentioned study of the theory revealed that some definite advantages could be derived by utilizing the average gate voltage directly as the output quantity. One obvious advantage, of course, results from the fact that this "output voltage" is thereby made independent of the regulation of the power-supply voltage. More important advantages, however, can be realized when the gate voltage is used to drive a subsequent stage of a multistage amplifier. It is shown later that the application of shunt coupling to a 2-stage magnetic amplifier effectively circumvents the traditional interstage coupling problems and yields an amplifier with all the advantages and few of the disadvantages of conventional 2-stage amplifiers.

Since, strictly speaking, shunt-coupled circuits do not represent a new class of magnetic amplifiers, all of the general statements relating to response time, voltage gain, driving-point impedance, etc., which apply to conventional self-saturating circuits also apply to shunt-coupled circuits. Shunt coupling is simply a method of utilizing conventional magnetic-amplifier circuits which is characterized by high signal efficiency, good stability, and circuitry simplicity.

Gate-Voltage Coupling

The basically simple method for coupling the gate voltage to a load may best be illustrated by reference to the simple half-wave magnetic-amplifier circuit of Fig. 1. During the indicated gating half-cycle if the voltage drop in the gate-winding resistance is sufficiently small, the half-cycle average-load voltage $\bar{e}_L$ will equal

---

Paper 59-635, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department. Manuscript submitted September 29, 1958; made available for printing March 9, 1959.

R. M. HUBBARD is with the Boeing Airplane Company, Seattle, Wash.

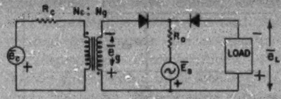


Fig. 1. Circuit diagram for the basic shunt-coupled magnetic amplifier

the half-cycle average-gate voltage $\bar{e}_g$. (Note that the diode-threshold-voltage drops tend to cancel.) The resistance R_0, which in a conventional circuit would be the load resistor, in the shunt-coupled circuit becomes merely a current-limiting resistor. Although the simple circuit of Fig. 1 is admittedly of limited practical value, merely exchanging efficiency for stability with changes in supply voltage and frequency, it forms the basis for several new circuits which represent distinct improvements over conventional circuits. It will be seen that the use of shunt coupling is most beneficial when the load is, in reality, a control winding on a separate core. This separate core may typically represent a subsequent stage of a multistage amplifier or it may be the complementary core in a full-wave amplifier.

A Solution to the Interstage Coupling Problem

Conventional techniques for coupling magnetic-amplifier stages in cascade are characterized by several deficiencies. For example, conventional full-wave stages in cascade, of necessity, employ power-wasting forcing-resistance to assure fast response in the second stage. Also, commonly, the highly peaked output-voltage waveform of the first stage leads to rectifier unblocking and associated problems in the second stage unless some interstage filtering is provided. This filtering, of course, lengthens the response time and adds to the complexity of the circuit. Ordinarily, the two stages are biased independently, a fact which does not contribute to good stability. Half-wave-type stages in cascade normally utilize more efficient coupling methods but they, as well as the full-wave circuits are sensitive to power-supply variations. Perhaps the most important limitation on the design of half-wave stages in cascade is imposed by the fact that the impedance presented by the control winding of the second stage must be small compared to the unsaturated impedance (excitation impedance) of the gate winding of the first stage. As has been indicated, shunt-coupling techniques effectively circumvent all of the afore-mentioned problems with the added bonus that the second stage, and any subsequent stage, of a multistage magnetic amplifier always inherently exhibits half-cycle response.

Consider the 2-stage half-wave shunt-coupled magnetic amplifier of Fig. 2(A), in which the two component saturable transformers are represented in the equivalent circuit form introduced by this author in an earlier paper.[1] As far as the equivalent circuit is concerned, it is sufficient to say that the nonlinear excitation impedance of each ideal saturable transformer is represented as a half-cycle average resistance $\bar{R}$ while the control-winding resistance and the gate-winding resistance are designated r_c and r_g respectively. All voltages are represented in terms of half-cycle average values and the polarities shown apply during the indicated half-cycle.

Good design practice dictates that the winding resistances r_{g1} and r_{c2} be negligibly small, with the result that the interstage coupling efficiency approaches 100% and the second-stage reset-voltage $\bar{e}_{r2}'$ is substantially equal to the first-stage gate voltage $\bar{e}_{g1}$. More specifically, under ideal conditions,

$$r_{g1}/R_0 \ll 1$$

$$r_{g1}/\bar{R}_2(N_{c2}/N_{g2})^2 \ll 1$$

$$r_{c2}/\bar{R}_2(N_{c2}/N_{g2})^2 \ll 1$$

and, as in the case of any conventional circuit,

$$R_0/\bar{R}_1 \ll 1$$

Although the foregoing assumptions greatly simplify the subsequent discussion of the shunt-coupled circuit, they are by no means essential to the good operation of the circuit. A more rigorous analysis, based on the afore-mentioned block-diagram techniques, in which the effects of winding resistance are included, is presented in the Appendix.

The fact that the second-stage reset voltage $\bar{e}_{r2}'$ is always substantially equal to the first-stage gate voltage $\bar{e}_{g1}$ means in effect that the operation of the second stage approaches that ideal condition associated with operation from a true voltage source. This type of operation is characterized by an over-all transfer characteristic which is substantially independent of nonlinearities in the excitation impedance of the second-stage core ($\bar{R}_2$). In effect, the second stage merely faithfully reproduces the first-stage output at a higher power level with an added half-cycle delay. The isolation provided by the blocking of diode D_2 during the second-stage gating half-cycle eliminates the gate-circuit loading which would ordinarily characterize operation from a low-impedance signal source. The presence of this same loading problem in the first stage, a Logan-type circuit, necessitates that the reflected control-circuit resistance be at least as large as the excitation impedance, that is

$$R_c(N_{g1}/N_{c1})^2 \geq \bar{R}_1$$

It is of interest to note here that the fact that the control-circuit resistance R_c must be appreciable to eliminate loading during the gating half-cycle is directly responsible for much of the nonlinearity introduced during the first-stage reset half-cycle. That is, if R_c could be negligibly small, its *I-R* drop caused by the finite nonlinear

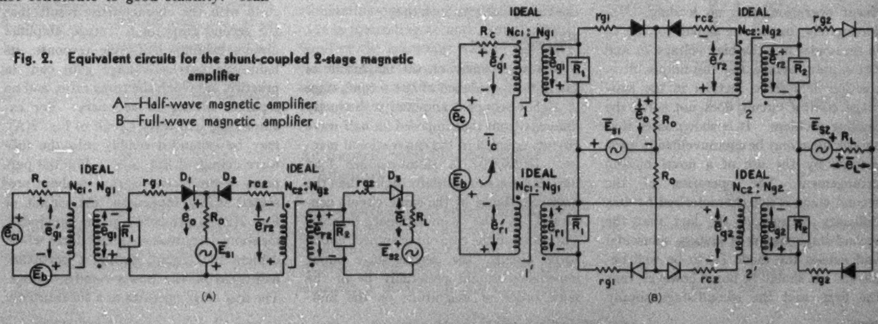
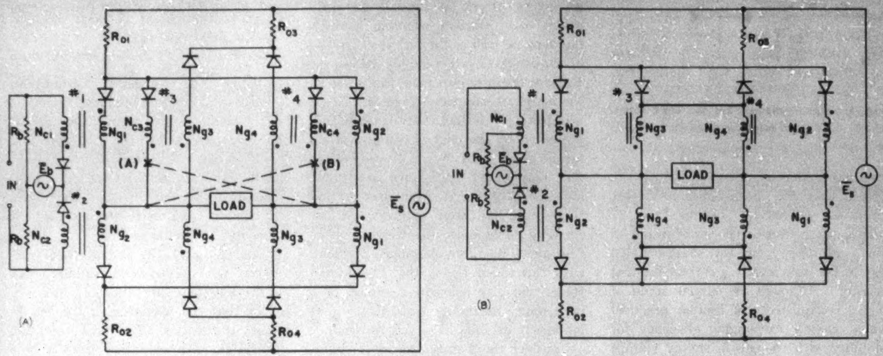


Fig. 2. Equivalent circuits for the shunt-coupled 2-stage magnetic amplifier

A—Half-wave magnetic amplifier
B—Full-wave magnetic amplifier



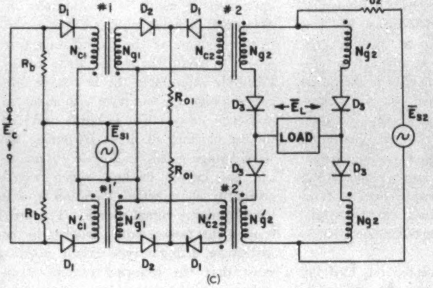


Fig. 3. **Circuit diagrams for shunt-coupled amplifiers**

A—Universal full-wave amplifier

B—Full-wave a-c amplifier

C—Differential half-wave amplifier

excitation impedance $\bar{R}_1$ would be negligible and the first-stage gain would be linear and equal to the turns ratio (N_{g1}/N_{c1}). Thus, if a synchronous rectifier were to be added to the first-stage control circuit in the same manner that the rectifier D_2 operates in the second-stage control circuit, the resistance R_c could be made negligibly small compared to the input impedance, $\bar{R}_1/(N_{c1}/N_{g1})^2$. The resultant operation would be characterized by linear operation with no loading. Unfortunately, unless the signal voltage $\bar{e}_{c1}$ is sinusoidal like the bias voltage $\bar{E}_b$ and the supply voltage $\bar{E}_{s1}$, the simple introduction of a series rectifier in the first-stage control circuit does not solve the loading problem. It is shown later that this problem may be circumvented in some circuits by the use of a novel biasing arrangement. The operation of the second stage is not characterized by this difficulty because of the fact that the second-stage signal is always sinusoidal and causes the diode D_2 to be well behaved. It should be noted however, that the first- and the second-stage supply voltages must be properly proportioned to assure that diodes D_2 and D_3 operate properly. That is, during the second-stage reset half-cycle, the instantaneous value of $\bar{E}_{s2}$ must always exceed the instantaneous value of the reset voltage $\bar{e}_{r2}$ to keep D_3 blocked. Likewise, during the gating half-cycle, the instantaneous value of $\bar{E}_{s1}$ must exceed the instantaneous value of $\bar{e}_{r2}'$ to keep D_2 blocked. This condition is readily obtained by selecting the supply voltages such that

$$\bar{E}_{s2}/\bar{E}_{s1} \simeq N_{g2}/N_{c2}$$

The requirement on the magnitude of the input impedance of the second stage of a shunt-coupled amplifier is distinctly different from that imposed on half-wave circuits coupled in the conventional manner. Ordinarily, a shunt-coupled load impedance is substantially higher than the load impedance of the same circuit connected in the conventional manner. Therefore, in the case of the subject multistage circuit, the second-stage input impedance should preferably be of the same order of magnitude as the first-stage excitation impedance. That is,

$$\bar{R}_2(N_{c2}/N_{g2})^2 \approx \bar{R}_1$$

This is in marked contrast to the requirement on the conventional circuit that the second-stage input impedance be at least an order of magnitude smaller than the first-stage excitation impedance.

Therefore, it can be seen that the operation of the second stage of a shunt-coupled 2-stage circuit approaches that theoretical ideal limit which is normally so difficult to achieve in practice. That is, since rectifier unblocking during reset and loading during gating are substantially nonexistent, the half-cycle average voltage gain can be made substantially equal to the turns ratio. Further, the extremely close coupling between stages obviates the necessity for separate bias on the second stage thereby eliminating a source of stability and loading problems.

Representative Shunt-Coupled Circuits

Shunt coupling may be applied to all the conventional magnetic-amplifier circuits with the characteristic result that the second stage of a 2-stage amplifier always exhibits half-cycle response, its half-cycle average-voltage gain can, in practice, approach the turns ratio, and no separate biasing is necessary. For example, the half-wave circuit of Fig. 2(A) may be expanded readily into the full-wave circuit of Fig. 2(B). For the purposes of the block-diagram analysis presented in the Appendix, the circuit of Fig. 2(B) is presented in the previously mentioned equivalent circuit form which depicts the circuit during a particular half-cycle of the power-supply voltage. The first stage operates as a substantially

Vol.
Pt. 1
195

conventional full-wave doubler circuit with the exception that a split load (R_0) is used to eliminate the rectifier-unblocking-problem characteristic of the doubler circuit with a single load. Thus, characteristically, the first stage has slow response and low input impedance, is capable of high-voltage gain, has a linearity characteristic directly related to the nonlinearity of the excitation impedance $\bar{R}_1$, and in all other respects operates as a conventional full-wave circuit. However, the shunt-coupled second stage is characterized by somewhat different operating characteristics. The isolation between the control windings provided by shunt coupling eliminates the intrinsic, positive, voltage feedback[1] which is characteristic of series-connected control windings, with the net result that the second stage exhibits half-cycle response. Further, the second stage input impedance is simply the reflected excitation impedance and the average voltage gain is simply the turns ratio (N_{g2}/N_{c2}), as was the case, of course, with the previously described half-wave circuit. Possible variations of this circuit would involve such things as using a center-tapped circuit in the first stage to eliminate the need for two limiting resistors, or the use of that type of circuit in the second stage to obtain d-c output.

Shunt-coupling techniques applied to a particular type of 2-stage half-wave amplifier, namely one in which the outputs of both stages are delivered to a common load, yield an extremely versatile full-wave a-c or d-c output "universal" amplifier. The basic concept involved is functionally similar to that contained in the full-wave circuits developed by Lord[3] and by Barnhart[4] but the method of implementation is appreciably different.

Table I. Parts List for the Experimental 2-Stage Amplifier of Fig. 4(B)

D_1	HD 6007 silicon diode, 55 ma (milliampere) 150 volts
D_2	HD 6006 silicon diode, 75 ma, 70 volts
D_3	Automatic 1N440 silicon diode, 300 ma, 100 volts
E_{s1}	10 volts rms, 400 cps (cycles per second)
E_{s2}	57.5 volts rms, 400 cps
N_{c1}	750 turns, no. 36 HF*
N_{c2}	110 turns, no. 32 HF
N_{g1}	900 turns, no. 32 HF
N_{g2}	550 turns, no. 28 HF
R_b	10K ohms, 1/2 watt
R_{s1}	500 ohms, 1/2 watt
R_{s2}	200 ohms, 4 watts
Load	200 ohms
First-stage cores	0.625×0.75×0.25×1 mil Hymu 80
Second-stage cores	1.0×1.5×0.375×1 mil, Hymu 80

* Heavy Formvar

Consider the circuit diagram of Fig. 3(A) in which cores 1 and 2 are basic to the first stage, a conventional half-wave bridge circuit. Similar cores 3 and 4 constitute a shunt-coupled second stage which delivers an output to the common load during the first-stage reset half-cycle. Note that in order to produce a balanced full-wave output, the second stage must have unity voltage gain and therefore all windings on all cores must be identical, with the exception of N_{c1} and N_{c2}. For illustration purposes the circuit is biased in a manner which effectively circumvents the previously mentioned loading problem during the gating half-cycle. The voltage E_b operates the two rectifiers in the control circuit in synchronism with the power supply in order to isolate the input circuit during the gating half-cycle. The indicated circuit may be operated from a control-voltage source having effectively zero internal impedance. This novel bias circuit is functionally the same as one previously described by Schohan.[5] The configuration shown in Fig. 3(A) provides a reversible-phase full-wave a-c output, but the simple interchange of the connections at points *A* and *B* as indicated by the dotted lines changes the circuit to yield a reversible-polarity full-wave d-c output.

The circuit of Fig. 3(A) may be simplified appreciably by restricting its operation to that of a reversible-phase a-c output amplifier. Such a simplified circuit diagram appears in Fig. 3(B). Note that both gating and reset are achieved in the same windings on cores 3 and 4 and that one pair of rectifiers is necessary for resetting and one pair for gating. Note that for illustration purposes the bias circuit has been slightly modified to a configuration which produces less loading of the control-voltage source. The rectifiers are still effective in isolating the control circuit during gating. A possible further simplification of Figs. 3(A) and 3(B) could be achieved by using an isolated power supply for each stage and thus eliminating one pair of limiting resistors.

It is interesting to note that the passband of the preceding class of full-wave circuit is the average of the passband of a single-stage amplifier and that of a 2-stage amplifier.

An Experimental Shunt-Coupled Amplifier

The application of shunt coupling to a 2-stage differential half-wave type of magnetic amplifier is illustrated by the circuit diagram of Fig. 3(C). The first stage is basically a differential half-wave circuit with a split load while the second stage is a shunt-coupled half-wave bridge circuit. The first-stage bias is applied in the previously described manner which affords the same type of decoupling during the gating half-cycle as is inherently provided by shunt coupling. Two resistors R_b provide for bias adjustment and zero balance for the whole amplifier, since shunt coupling obviates the necessity for separately biasing the second stage.

The parts list and other specifications for the circuit of Fig. 3(C) appear in Table I and the experimental operation of the circuit is summarized by the results presented in Figs. 4(A), 4(B), and 5. The transfer curve of Fig. 4(A) gives a plot of the fundamental-frequency output voltage versus control voltage. The control-voltage source was a Variac which presented substantially zero internal impedance to the amplifier input terminals.

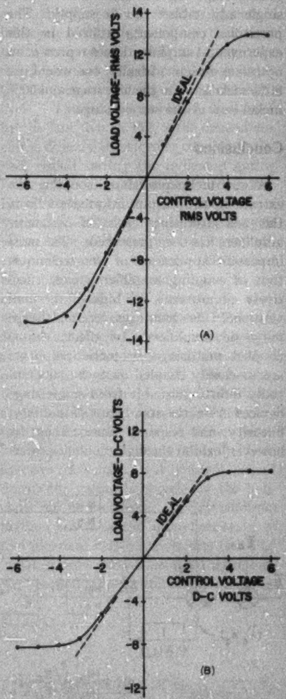


Fig. 4. Transfer curves for the experimental amplifier of Fig. 3(C)

A—A-c curves
B—D-c curves

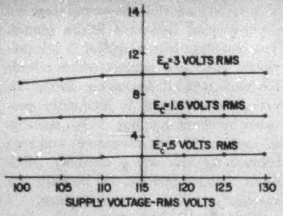


Fig. 5. Effect of power-supply regulation on the experimental amplifier of Fig. 3(C)

The transfer curve of Fig. 4(B) was obtained with a pure d-c control-voltage source also having substantially zero internal impedance. Fig. 5 illustrates the good stability of this circuit under conditions of changing power-supply voltage. Both the 10-volt supply for the first stage and the 57.5-volt supply for the second stage were transformed from the single adjustable 115-volt supply. The particular components utilized in this experimental amplifier do not represent an optimum design. Ideally, one would use 80% nickel-iron in the first stage and 50% nickel-iron in the second stage.

Conclusions

A new technique, shunt coupling, for extracting the output information from the self-saturating class of magnetic amplifiers has been presented. The most important application of this technique, that of coupling amplifier stages, effectively circumvents the bulk of the conventional disadvantages of multistage magnetic amplifiers. In effect, shunt-coupled multistage magnetic amplifiers are so closely coupled as to be substantially indistinguishable from single-stage devices from the standpoint of stability, linearity, and response time. They do, however, exhibit the high-dynamic power-gain characteristic of multistage amplifiers. This close coupling technique applied to 2-stage half-wave magnetic amplifiers, in which both stages operate into a common load, has been shown to yield an extremely simple, full-wave, a-c or d-c output universal amplifier.

It is hoped that this introduction of a basically simple new coupling concept will stimulate the re-examination of other conventional magnetic-amplifier circuits for the purpose of simplification or improvement in operating characteristics.

Appendix

Development of the Block Diagram

The circuit of Fig. 2(B) may be analyzed, without the necessity of employing the restrictive assumption that the winding resistances should be negligible, by making use of the block-diagram method of analysis.[1] Further, it is felt that the block-diagram method of analysis provides an insight into the mechanics of operation of the subject circuit type which is difficult to achieve by studying the circuit diagram alone.

The block-diagram method of analysis is based on the recognition that magnetic-amplifier operation can be described in terms of half-cycle average values and that conventional methods of circuit analysis (Kirchoff's laws, etc.) are usable when all voltages and currents are so expressed. Further, it must be recognized that the average voltage across a given winding over a complete cycle of resetting and subsequent gating must be zero and, therefore, if commutation effects are negligible, the magnitude of the half-cycle average voltage during reset must equal the magnitude of the half-cycle average voltage during gating. This fact may be expressed in Laplace notation as:

$$|\bar{e}_g(s)| = \epsilon^{-Ts}|\bar{e}_r(s)| \tag{1}$$

In the foregoing $\bar{e}_g(s)$ is the half-cycle average gate voltage, $\bar{e}_r(s)$ is the half-cycle average reset voltage, and T is time for one half-cycle of the power-supply voltage.

Write the loop equations for the circuit of Fig. 3 starting with the control circuit of the first stage and dropping the argument (s) for simplicity. During the indicated half-cycle

$$\bar{E}_b - \bar{e}_c + \bar{e}_{g1}' - \bar{i}_c R_c = \bar{e}_{r1}' \tag{2}$$

Since the rectifier associated with the resetting core $1'$ is blocked during the subject half-cycle, it can be seen that

$$\bar{i}_c = \bar{e}_{r1}'/\bar{R}_1(N_{c1}/N_{g1})^2 \tag{3}$$

For simplicity the reflected excitation impedance is designated as

$$\bar{R}_1(N_{c1}/N_{g1})^2 = \bar{R}_1' \tag{4}$$

Using equations 3 and 4 it can be seen that

$$\bar{i}_c R_c = \bar{e}_{r1}'[R_c/\bar{R}_1'] \tag{5}$$

Substituting equation 5 into 2 and combining terms yields

$$[\bar{E}_b - \bar{e}_c + \bar{e}_{g1}']\left[\frac{\bar{R}_1'}{\bar{R}_1' + R_c}\right] = \bar{e}_{r1}' \tag{6}$$

For simplicity of notation, let the first-stage intrinsic amplification factor be given by

$$A_{i1} = \frac{\bar{R}_1'}{R_c} = \frac{(N_{c1})^2/R_c}{(N_{g1})^2/\bar{R}_1} \tag{7}$$

Substituting equation 7 into 6 yields

$$[\bar{E}_b - \bar{e}_c + \bar{e}_{g1}']\left[\frac{1}{1+(1/A_{i1})}\right] = \bar{e}_{r1}' \tag{8}$$

Using equation 1 and noting that

$$\bar{e}_{r1} = \bar{e}_{r1}'(N_{g1}/N_{c1}) \tag{9}$$

it follows that

$$\bar{e}_{g1} = \epsilon^{-Ts}\bar{e}_{r1} = \epsilon^{-Ts}[\bar{e}_{r1}'(N_{g1}/N_{c1})] \tag{10}$$

Also,

$$\bar{e}_{g1}' = \bar{e}_{g1}(N_{c1}/N_{g1}) \tag{11}$$

Clearly, equations 8, 10, and 11 contain sufficient information to permit the construction of the block diagram for the first stage as shown in Fig. 6.

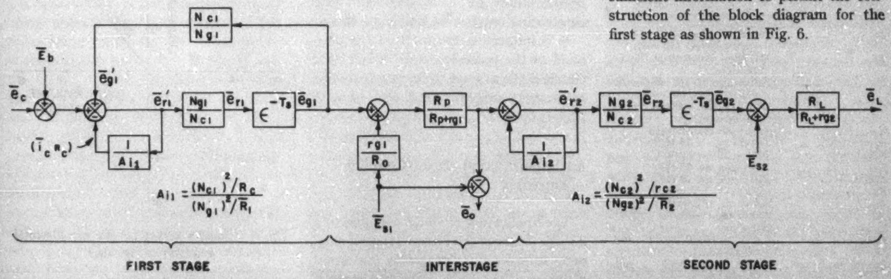


Fig. 6. Complete block diagram for the 2-stage full-wave amplifier of Fig. 2(B)

The equations for the operative interstage network associated with gating core 1 and resetting core 2 may be written directly from the equivalent circuit of Fig. 2(B) since, due to symmetry, the resultant equations are identical with those which would be written for cores 1′ and 2′ during the subsequent half-cycle. Thus, if the two loop currents are designated $\bar{i}_1$ and $\bar{i}_2$ and the reflected excitation impedance of core 2 is designated $\bar{R}_2'$, the two mesh equations may be written as follows:

$$\bar{i}_1(r_{g1}+R_0)+\bar{i}_2(R_0)=\bar{E}_{s1}-\bar{e}_{g1} \tag{12}$$

$$\bar{i}_1(R_0)+\bar{i}_2[R_0+r_{c2}+\bar{R}_2']=\bar{E}_{s1} \tag{13}$$

The simultaneous solution of the foregoing equations and the subsequent substitution of

$$\bar{e}_{r2}'=\bar{i}_2\bar{R}_2' \tag{14}$$

and

$$\bar{e}_0=(\bar{i}_1+\bar{i}_2)R_0 \tag{15}$$

yields

$$[\bar{e}_{g1}+\bar{E}_{s1}(r_{g1}/R_0)]\times\left[\frac{1}{1+r_{g1}[1/(R_0)+1/(r_{c2}+\bar{R}_2')]}\right]\times\left[\frac{\bar{R}_2'}{\bar{R}_2'+r_{c2}}\right]=\bar{e}_{r2}' \tag{16}$$

Also

$$[\bar{E}_{s1}-\bar{e}_0]\left[\frac{\bar{R}_2'}{\bar{R}_2'+r_{c2}}\right]=\bar{e}_{r2}' \tag{17}$$

or

$$\bar{E}_{s1}-\bar{e}_{r2}'\left[\frac{\bar{R}_2'+r_{c2}}{\bar{R}_2'}\right]=\bar{e}_0 \tag{18}$$

The foregoing expressions may be simplified by letting the gross load resistance seen by the first stage be designated R_p. Thus,

$$\frac{1}{R_p}=\frac{1}{R_0}+\frac{1}{r_{c2}+\bar{R}_2'} \tag{19}$$

Also the intrinsic amplification factor of the second-stage control circuit is given by

$$A_{i2}=\frac{\bar{R}}{r_{c2}}=\frac{(N_{c2})^2/r_{c2}}{(N_{g2})^2/\bar{R}_2} \tag{20}$$

Substituting equations 19 and 20 into 16 and 18 yields

$$[\bar{e}_{g1}+E_{s1}(r_{g1}/R_0)][R_p/(R_p+r_{g1})]\left[\frac{1}{1+(1/A_{i2})}\right]=\bar{e}_{r2}' \tag{21}$$

and

$$\bar{E}_{s1}-\bar{e}_{r2}'[1+(1/A_{i2})]=\bar{e}_0 \tag{22}$$

Equations 21 and 22 permit the construction of the block diagram for the interstage network and for the control circuit of the second stage.

Note that with relation to resetting core 2,

$$\bar{e}_{r2}'(N_{g2}/N_{c2})=\bar{e}_{r2} \tag{23}$$

and also from equation 1

$$\bar{e}_{g2}=\epsilon^{-T_s}\bar{e}_{r2} \tag{24}$$

which yields

$$\bar{e}_{g2}=\epsilon^{-T_s}[\bar{e}_{r2}'(N_{g2}/N_{c2})] \tag{25}$$

In the load circuit of the gating core 2′

$$[\bar{E}_{s2}-\bar{e}_{g2}]\left[\frac{R_L}{R_L+r_{g2}}\right]=\bar{e}_L \tag{26}$$

Equations 25 and 26 permit the completion of the block diagram for the second stage as shown in Fig. 6.

The symmetry of the full-wave circuit makes all of the foregoing equations valid for that half-cycle which was not explicitly treated in the equivalent circuit.

It can be seen from equations 16 and 18 that if

$$r_{g1}\ll R_0 \tag{27}$$

$$r_{g1}\ll \bar{R}_2' \tag{28}$$

and

$$r_{c2}\ll \bar{R}_2' \tag{29}$$

then the subject equations reduce respectively to

$$\bar{e}_{g1}=\bar{e}_{r2}' \tag{30}$$

and

$$E_s-\bar{e}_{r2}'=\bar{e}_0 \tag{31}$$

The resultant simplified block diagram appears in Fig. 7. This simplified block diagram is applicable to the introductory discussion which appears in the text.

Observations Based on the Half-Cycle Average Method of Analysis

In the author's opinion, the half-cycle average method of analysis in general and the block-diagram method of analysis in particular provide a fresh approach to the explanation of familiar phenomena associated with magnetic-amplifier circuitry. At the risk of excessive digression from the main topic of this paper, an effort is made herein to exploit the novel features of the subject methods of analysis in order to stimulate a re-examination of some of the classical methods of explaining magnetic-amplifier operation. It is suggested that reference be made both to the equivalent circuit of Fig 2(B) and to the block diagram of Fig. 6, while following the subsequent discussion.

The simple half-wave magnetic amplifier, as typified by the first stage of the circuit in Fig. 2(A), is the basic building block of which all conventional self-saturating amplifiers are constructed. Ideally, this simple amplifier can accept an input signal only during one (reset) half-cycle and can deliver an output only during the subsequent (gating) half-cycle of the power-supply voltage. The (shunt) input impedence during the reset half-cycle is simply the reflected excitation impedance $\bar{R}_1'$ and the response time in terms of half-cycle average values is one half-cycle pure delay of duration T. Aside from the fact that the maximum attainable output corresponds to only a half-cycle of the supply voltage, the simple half-wave circuit suffers from the deficiency that the control circuit loads the gate circuit during the gating half-cycle. This loading problem is illustrated by the fact that in Fig. 2 the control circuit is directly transformer coupled to the gate circuit during the indicated gating half-cycle. Thus the control circuit does in fact have an effect on the operation of the amplifier during the gating half-cycle rather than solely during the reset half-cycle. In order to minimize this loading effect, several courses are open to the circuit designer. For example, he may adjust the resistance in the control circuit R_c and/or the turns ratio to reflect a sufficiently high impedance from the control circuit into the gate circuit. Or, he may use a series inductor to raise the impedance of the control circuit without fear of affecting the response time. Another possibility, which is mentioned in the text, involves decoupling the control circuit during the gating half-cycle by the judicious use of rectifiers. A very common method involves the interconnection of two identical half-wave circuits (see Fig. 2(B)) in order to deliver a full-wave output as well as to eliminate loading by

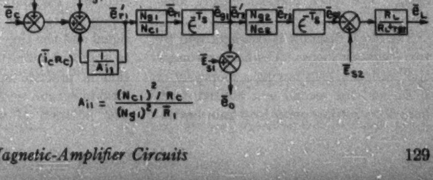


Fig. 7. Simplified block diagram for the circuit of Fig. 2(B) when the winding resistances are negligible

virtue of series-connected control windings. The simple act of series connecting the control windings of two simple half-wave amplifiers has such a profound effect on the operation of the amplifiers that a detailed study is warranted.

As a result of a holdover from the study of simple saturable-reactor circuits, the series connection of control windings in self-saturating circuits is often described as a phase-opposition connection which causes the induced voltages to buck. In reality, only the gating core is characterized by an induced voltage ($\bar{e}_{g1}'$). The resetting core reflects the necessary high impedance into the control circuit to eliminate loading in a manner somewhat analogous to the insertion of a series inductor in the control circuit of the simple half-wave circuit. The reflected impedance $\bar{R}'_1$ of the resetting core limits the average current in the control-circuit ($\bar{i}_c$) to the comparatively low excitation current. Of course, in some cases the resistance in the control circuit is of itself sufficient to eliminate loading, that is,

$$R_c/\bar{R}_1'>1 \text{ or } A_{fl}<1$$

This latter condition is often referred to as suppressed even-harmonic operation. Those former cases in which the control-circuit resistance R_c is small compared to $\bar{R}_1'$ are referred to as free even-harmonic operation.

It is clear that the fraction of the total voltage introduced into the control circuit which actually is effective in resetting the core is determined by the voltage division between R_c and $\bar{R}_1'$. This fact is clearly illustrated by equation 6. It is convenient to treat this voltage drop across R_c as a negative feedback voltage in the manner suggested by equations 5, 7, and 8, and as illustrated by the block diagram.

In the block diagram, the full-wave circuit is treated as an entity and the individual cores lose their separate identities over a complete cycle. Thus, during each half-cycle, one core is always resetting and one core is always gating and the circuit equations are independent of the particular half-cycle selected for study. Since the magnitude of the induced gate voltage $\bar{e}_{g1}'$ during the indicated half-cycle is equal to the magnitude that the reset voltage $\bar{e}_{r1}'$ had during the preceding half-cycle, it is clear that $\bar{e}_{g1}'$ is in reality a feedback voltage delayed by one half-cycle. The block diagram clearly illustrates this fact.

The block diagram is characterized by two feedback loops. The first is a negative feedback loop having a loop gain of $1/A_{fl}$ and the second is a positive feedback loop having unity loop gain but incorporating a pure delay. It is a simple matter to draw some general conclusions about the full-wave amplifier at this point. Note that the input voltage to the basic amplifier is the reset voltage $\bar{e}_{r1}'$ and the input impedance is the reflected excitation impedance $\bar{R}_1'$. The feedback voltage $\bar{e}_{g1}'$ introduced by the control winding of the gating core is directly in series with the input and serves to modify the impedance presented by the amplifier to the external signal source. In particular, the unity gain, positive, voltage-feedback loop lowers the effective input impedance to zero in the steady state. This explains the well-known fact that, in the steady state, the current in the control circuit is limited solely by the d-c resistance R_c and the amplifier can be described as current sensitive (or ampere-turns sensitive). This is clarified by noting that, in the steady state, the gate voltage during one half-cycle must equal the gate voltage during the subsequent half-cycle. Thus, during a given half-cycle, the gate voltage $\bar{e}_{g1}'$ must equal the reset voltage $\bar{e}_{r1}'$. It is therefore clear from the equivalent circuit and from the block diagram that a necessary condition or a steady state is

$$\bar{E}_b-\bar{e}_c=\bar{i}_cR_c$$

Although the magnetic amplifier is apparently current-sensitive in the steady state, it is most certainly voltage sensitive during transients because of the fact that the positive feedback is delayed. If this were not the case, it would be impossible to develop a voltage across the input terminals and thereby to achieve a change in reset voltage.

The negative feedback loop, which is not delayed, also has an important effect on the operation of the magnetic amplifier. In the absence of the subject loop, the unity gain positive feedback loop would cause the voltage gain of the magnetic amplifier to approach infinity. That is, if R_c went to zero, then $1/A_{fl}$ would go to zero and a steady state could not be achieved with any finite value of input voltage. Thus, the negative feedback loop lowers the voltage gain and provides stability.

It is of interest to investigate the source of the familiar magnetic-amplifier time constant by inspection of the block diagram. In the absence of negative feedback, the amplifier cannot achieve a steady state after the application of a signal and the time constant may be said to approach infinity. In fact, the operation in practice approaches that of an ideal integrator with an initial time delay. The addition of negative feedback by making R_c finite means that a steady state will ultimately be achieved. Thus a transient will ensue after the application of an input until the reset voltage has increased, by virtue of the positive feedback, to a level sufficient to provide the equilibrium negative feedback voltage. Note that if the negative feedback gain were unity, a steady state would be achieved in two half-cycles of duration T. If the negative feedback gain were to approach infinity, the positive feedback would be negated and the response time would be simply one half-cycle. This condition would obtain for very large values of R_c relative to $\bar{R}_1'$.

It should be apparent from the block diagram that, in the absence of the positive feedback loop, the response time would be simply one half-cycle (T) and the voltage gain could never exceed the turns ratio N_{g1}/N_{c1}. It also can be seen from the circuit diagram that, in the absence of the subject loop, the input impedance would be simply $\bar{R}_1'$. The elimination of the positive feedback loop is simply a matter of providing isolation between the two control windings. Shunt coupling does exactly this as evidenced by the second-stage block diagram. It is clear that the second stage is characterized by a half-cycle response time.

It can be seen from the block diagram of Fig. 6 that the voltage $\bar{e}_0$ across the limiting resistor R_0 has no direct effect on the second stage. However, the circuit is not completely insensitive to variations in the power supply $\bar{E}_{s1}$ unless the first-stage gate-winding resistance r_{g1} is small compared to the limiting resistor R_0. Further, the efficiency of the interstage is limited by the ratio of r_{g1} to the effective load on the first stage R_p. The control circuit efficiency of the second stage is limited by the ratio of the control-winding resistance r_{c1} to the reflected excitation impedance R_2'. Equations 27, 28, and 29 summarize the necessary conditions to make the coupling efficiency approach 100% and to cause the block diagram to simplify to that of Fig. 7.

Nomenclature

A_f=amplification factor
D=designates rectifying element
$\bar{E}_b$=average steady-state bias voltage
$\bar{e}_c$=half-cycle average control voltage
$\bar{e}_g$=half-cycle average gate voltage across gate winding
$\bar{e}_g'$=half-cycle average gate voltage referred to control winding
$\bar{e}_L$=half-cycle average load voltage (across R_L)
$\bar{e}_0$=half-cycle average voltage across current-limiting resistor R_0
$\bar{e}_r$=half-cycle average reset voltage across gate winding

$\bar{e}'_r$ = half-cycle average reset voltage referred to control winding
$\bar{E}_s$ = average rectified supply voltage
$\bar{I}_c$ = half-cycle average current in control winding
N_c = number of turns in control winding
N_g = number of turns in gate winding
$\bar{R}$ = average excitation impedance of saturable transformer referred to the gate winding
$\bar{R}'$ = average excitation impedance of saturable transformer referred to the control winding
R_b = total series resistance in bias circuit
R_c = total series resistance in control circuit
r_c = resistance of control (N_c) windings
r_g = resistance of gate winding
R_L = load resistance
R_o = current-limiting resistance
R_p = effective load resistance on first stage
s = complex frequency variable
T = time for one half-cycle of power-supply voltage
ϵ = base of natural logarithms

References

1. MAGNETIC AMPLIFIER ANALYSIS AND APPLICATION USING BLOCK-DIAGRAM TECHNIQUES, R. M. Hubbard. *AIEE Transactions*, vol. 76, pt. I, Nov. 1957, pp. 578–88.

2. THE TRANSACTOR: A SELF-SATURATED TRANSFORMER—PART I, Allen B. Rosenstein. *Ibid.*, vol. 77, pt. I, May 1958, pp. 129–42.

3. MAGNETIC AMPLIFIER CIRCUITS WITH FULL-WAVE OUTPUT AND HALF-WAVE CONTROL SIGNALS, Harold W. Lord. *Ibid.*, vol. 73, pt. I, July 1954, pp. 265–70.

4. A NEW FULL-WAVE MAGNETIC AMPLIFIER OUTPUT STAGE, P. W. Barnhart. *Ibid.*, vol. 74, pt. I, Mar. 1955, pp. 139–41.

5. A TRANSIENT-CONTROLLED MAGNETIC AMPLIFIER, George Schohan. *NAVORD Report 4258*, Naval Ordnance Laboratory, Washington, D. C., March, 1956.

Coolant Instrumentation for Pressurized Water Reactor Power Plants

J. J. LIHOTA
MEMBER AIEE

C. B. HOPPA
MEMBER AIEE

NUCLEAR POWER has become a reality. The theory and classical experiments of half a century have been assembled and within the last decade have provided mankind with a new source of power. The problem of handling and controlling this new process is a monument in itself to technical ingenuity of the American engineer. At this stage of the art one cannot judge the nuclear power plant as a simple system, but with new technique and every day improvements this new industry will be established on an economical basis.

This paper deals with the instrumentation required of the medium cooling the reactor, and the facilities rendered by this instrumentation for control, indication, and safety.

Reactor Coolant Cycle

There are many processes that can be utilized for transforming the energy of the atom. For the purpose of this paper discussion will be limited to the pressurized water type of reactor system. A typical nuclear power plant main coolant system is shown in Fig. 1. In the interest of simplicity only one coolant loop is detailed; however, each of the number of loops connected would be physically and functionally identical.

The heat generated by the fuel in the reactor is transported by the coolant medium through the connecting piping to a heat exchanger wherein the energy is transferred to a second water system to be utilized as a steam source for the work load. It is necessary that a sufficient flow rate be maintained to prevent localized heating of the fuel elements. The main coolant pump returns the coolant water to the reactor for reheating. The system thus far discussed is a fluid water system.

An appendage to this system is connected to a pressurizer vessel wherein electric elements heat a relatively stagnant portion of the system water volume to a temperature and pressure sufficient to maintain the water volume in the circulating system in a fluid state. The steam blanket generated in the top of the pressurizer will also relieve the system of surges in water volume during operation.

From the foregoing it is obvious that pressure and temperature are two extremely important parameters in the design of a pressurized water reactor system. Likewise the accurate measurement of these two parameters becomes highly important in the safe operation of a pressurized water reactor system. The measurement of the temperature of the water as it leaves the reactor and as it enters the reactor is indicative of operating conditions within the reactor. The same applies to the measurement of system pressure.

Of a lesser degree of importance, but nevertheless important in the safe operation of the reactor power plant, is instrumentation for the following:

1. The pump head, in systems utilizing two speed pump control.
2. Coolant flow, in systems utilizing variable speed pump control.
3. Main loop stop valves position indication, either fully open or fully closed.
4. Pressurizer liquid level measurement.
5. Pressurizer temperature.

Modular Magnetic-Amplifier Instrumentation

An instrumentation system requires a parameter to be measured, a detector to transduce the physical change of the parameter into an electrical signal, and a receiver to utilize the signal for indication or control.

The detectors used to measure the plant parameters of fluid temperature, pressure, level, and flow are conventional. It is from the physical design aspects that these detectors become highly special. First the detectors must be designed so that all parts in contact with the fluid of the primary system are hermetically sealed either by welding or brazing. Second, the material used in the fabrication of the detectors must be specially selected. It must allow no matter to be introduced into the primary system which can become highly radioactive; and it must not become damaged by radiation. Third, because the detectors will be inaccessible for the extended periods of time that the reactor plant is critical, they must be designed for the highest degree of reliability and be relatively free of maintenance demands. Fourth, to provide the highest degree of plant protection the detectors must be designed to respond as rapidly as possible to changing plant conditions. Fifth, to provide for most efficient utilization of the reactor power

---

Paper 58-1257, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted August 20, 1958; made available for printing January 28, 1959.

J. J. LIHOTA and C. B. HOPPA are with the Westinghouse Electric Corporation, East Pittsburgh, Pa.

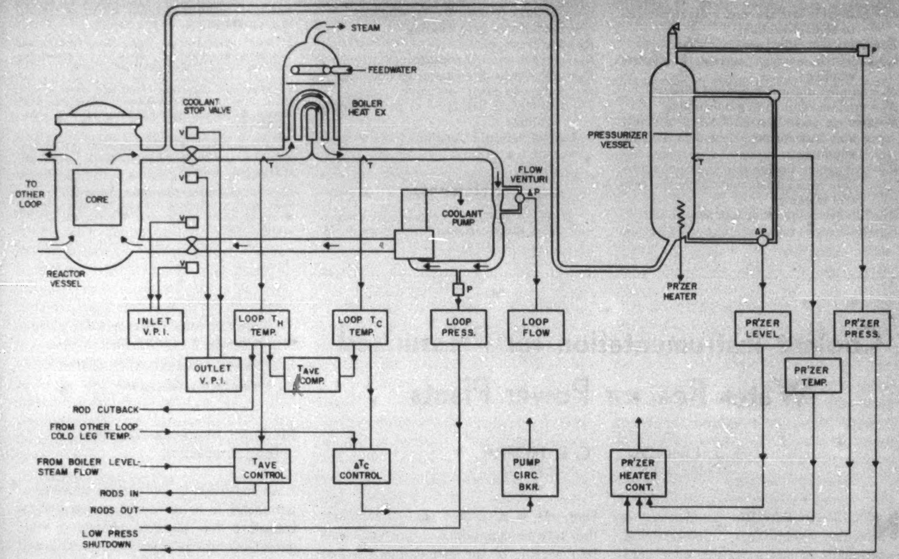
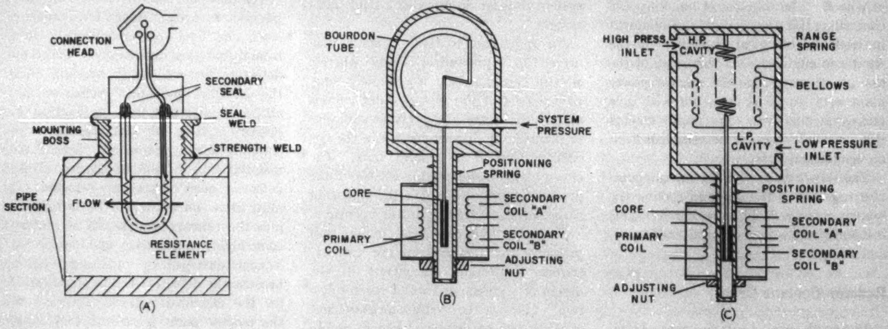

STEAM
FEEDWATER
BOILER HEAT EX
COOLANT STOP VALVE
TO OTHER LOOP
CORE
REACTOR VESSEL
V
T
FLOW VENTURI
ΔP
COOLANT PUMP
P
PRESSURIZER VESSEL
PR'ZER HEATER
INLET V.P.I.
OUTLET V.P.I.
LOOP T_H TEMP.
LOOP T_C TEMP.
T_AVE COMP.
LOOP PRESS.
LOOP FLOW
PR'ZER LEVEL
PR'ZER TEMP.
PR'ZER PRESS.
ROD CUTBACK
FROM OTHER LOOP COLD LEG TEMP.
FROM BOILER LEVEL-STEAM FLOW
RODS IN
RODS OUT
LOW PRESS SHUTDOWN
T_AVE CONTROL
ΔT_C CONTROL
PUMP CIRC. BKR.
PR'ZER HEATER CONT.

Fig. 1. Typical primary plant instrumentation system block diagram


CONNECTION HEAD
SECONDARY SEAL
SEAL WELD
MOUNTING BOSS
STRENGTH WELD
PIPE SECTION
FLOW
RESISTANCE ELEMENT
(A)
BOURDON TUBE
SYSTEM PRESSURE
POSITIONING SPRING
CORE
SECONDARY COIL "A"
PRIMARY COIL
SECONDARY COIL "B"
ADJUSTING NUT
(B)
HIGH PRESS. INLET
H.P. CAVITY
RANGE SPRING
BELLOWS
L.P. CAVITY
LOW PRESSURE INLET
POSITIONING SPRING
CORE
PRIMARY COIL
SECONDARY COIL "A"
SECONDARY COIL "B"
ADJUSTING NUT
(C)

A—Resistance thermometer installation

B—Pressure detector

C—Differential pressure detector

Fig. 2. Sectional views

source the detectors must have the highest accuracy and reproducibility.

In addition to the aforementioned design criteria the detectors for reactor coolant instrumentation have been designed to be especially suited for use with the magnetic-amplifier type of circuits.

Temperature is measured by means of a resistance thermometer element inserted into the fluid; see Fig. 2(A). The element electrically forms one leg of a Wheatstone bridge circuit so that as the resistance of the element varies with temperature the resultant bridge unbalance becomes a function of temperature.

Pressure is measured by a bourdon tube type of detector as in a conventional pressure gage, Fig. 2(B). The tube tip is connected by suitable linkage to a movable core of a differential transformer. The motion of the tube tip with pressure is transferred to the movable core, which in turn varies the coupling between the primary winding and two secondary windings of the differential transformer. When the outputs of the two secondary windings are electrically subtracted, the net result is an electrical output which is a function of pressure.

Liquid level and flow are measured by the differential pressure type of detectors; see Fig. 2(C). For liquid level measurements the detector measures the difference in head between a reference column of liquid and the level of the liquid in ques-

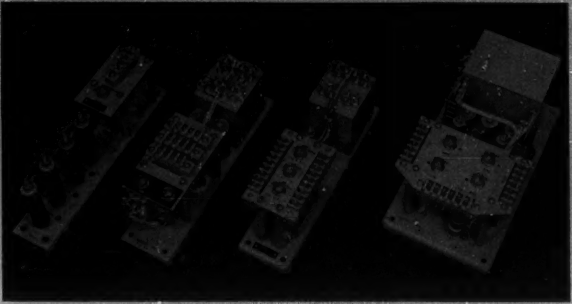


Fig. 3. Typical magnetic-amplifier modules (left to right, bridge, bistable controller, averaging amplifier, self-balancing amplifier)

tion. For flow measurements the detector measures the pressure drop across a venturi or an orifice plate. In either case the differential pressure acting on the bellows of the detector results in the motion of a movable core of a differential transformer. The position of the movable core is therefore a function of differential pressure. As in the pressure detector, for level measurements the movable core varies the coupling between the primary winding and secondary windings of a differential transformer, which results in an electrical output which is a function of level. In the case of flow measurements, because the pressure drop produced by the restriction is proportional to the flow squared, a specially designed differential transformer whose output is a square root function of head is used. As a result, for flow measurements the electrical output of the differential pressure flow detector is a linear function of flow.

The receivers vary in function but can be resolved to a few basic designs to fulfill most system requirements. In the heart of the reactor plant instrument, exists the magnetic amplifier. This device consists of windings on a core that exhibits rectangular hysteresis loop qualities and is inserted in the circuit between a power supply and the load to act as a variable impedance. The application of additional isolated windings and their proper excitation render the device as a controllable impedance with amplifying characteristics. The basic modules that are required for most instrumentation are: 1. bridge module, 2. demodulator module, 3. self-balancing amplifier module, 4. bistable controller module, and 5. summing amplifier module. A few of these typical modules are shown in Fig. 3.

The bridge module is used in temperature-measuring instruments to provide a voltage to the amplifier proportional to the temperature detected. The bridge is a Wheatstone circuit of which the detector resistive element constitutes one leg and which contains the necessary power supply and balance adjustments.

The demodulator module is applicable to detectors utilizing differential transformers as a signal source, rectifying the outputs of the secondaries, and determining the difference of their values to be presented to the amplifier as a single signal.

The amplifier module is a versatile circuit designed to utilize negative electric feedback for stability. Essentially it is a self-balancing magnetic amplifier which, because of the type of feedback employed, appears as a current source. The circuit possesses the advantage of eliminating drift, and nullifies the effects of frequency and voltage variation. Amplifiers of this type require a very low drain on the signal source during steady-state operation. The standard module is applicable as a wide-span or narrow-span amplifier and is adaptable to any of the previously discussed detectors by placing a bridge or demodulator module ahead of the amplifier. The output of this amplifier module is nominally zero to 20 milliamperes direct current. The accuracy for adverse conditions of voltage, frequency, and temperature is ±0.5%. The nonlinearity over the range is not measurable. The load regulation from 500 to 2,000 ohms is ±0.5%.

The bistable controller module (Fig. 4) is a self-saturating magnetic amplifier with positive feedback which provides a step change in output at a sharply defined and controllable setpoint. Control is provided to adjust the operating differential of this setpoint between input change limits of 1.0% to nominally 20% of span with extended ranges available. The set point is resolvable to within ±0.15% of input span and is accurate to within ±0.2% for adverse conditions of voltage, frequency, and temperature. Output for the tripped condition is nominally 320 milliamperes direct current at 28 volts.

The summing amplifier module is a simple saturable magnetic amplifier with

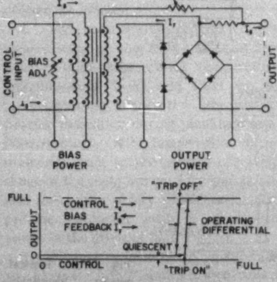


Fig. 4. Bistable controller

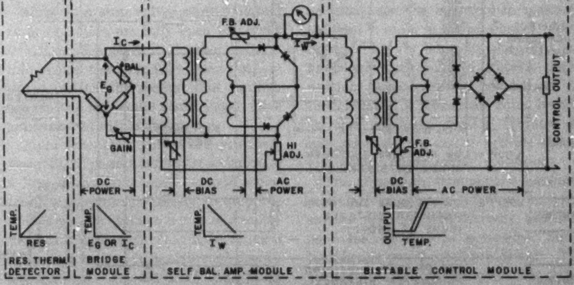


Fig. 5. Typical temperature instrument channel

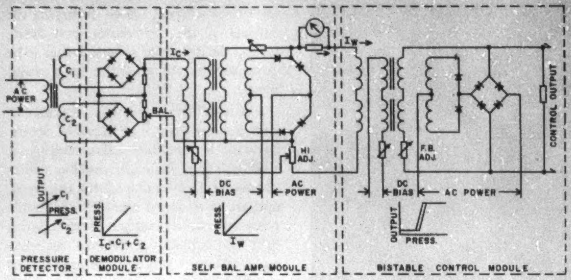

AC POWER
C1
C2
IC
BAL
IW
CONTROL OUTPUT
HI ADJ.
F.B. ADJ.
DC BIAS
AC POWER
DC BIAS
AC POWER
OUTPUT
C1
PRESS.
C2
PRESS.
IC=C1+C2
PRESS.
IW
OUTPUT
PRESS.
PRESSURE DETECTOR
DEMODULATOR MODULE
SELF BAL. AMP. MODULE
BISTABLE CONTROL MODULE

Fig. 6. Typical differential transformer instrument channel

multiple control windings. Proper wiring connection to the input of the module permits addition and subtraction operations magnetically with circuit isolation as well as power amplification. The characteristics and output are comparable to the self-balancing amplifier already mentioned.

It is worth noting that the modules described are of the same general construction, and that their input and output characteristics are selected to provide matching units assembled into complete instruments.

Fig. 5 shows a simple diagram for a temperature instrument, and the first application of the modules into a complete instrument. A resistance thermometer detector has its element connected to the bridge module. The bridge-balancing rheostat nulls the output voltage E_G at maximum temperature, providing the best accuracy in the operating span and providing an output inversely proportional to temperature. The output voltage is rather weak for control application and a self-balancing amplifier module is utilized to strengthen the control signal and apply it to the controllers. The single amplifier can provide either wide-span or narrow-span operation. Where it is desirable to have both, a second amplifier module is paralleled in the circuit fulfilling the requirements.

Instruments in applications where a differential transformer transduces the parameter measurement (pressure, level, and flow) utilize the modules outlined for the temperature instrument except that the bridge module is replaced by a demodulator module; see Fig. 6. Again, service can be provided for wide-span or narrow-span, and for wide- and narrow-span operation.

A typical computer channel provides for shutdown of the reactor for any one of a number of dangerous parameter conditions in the system. A master shutdown bistable module (Fig. 7) similar to the previously described bistable controller module, but with a larger output capacity, is utilized. An auctioneer module monitors the outputs of several shutdown parameter bistables in various measuring channels. The control and bias windings are connected circuitwise to maintain full output for no control signal in. When a dangerous parameter condition is detected, the parameter bistable controller will "turn on" delivering full output via the auctioneering diode to the master shutdown bistable control winding. A single control signal such as this will "turn off" the master shutdown bistable and its output will be reduced to a value near zero established by the quiescent value of its core. The output normally energizes a solenoid latch in the rod mechanism. In "turning off" the solenoid is de-energized and the control rods are quickly inserted into the reactor core.

The aforementioned computer is a simple decision-making device. A second type of computer continuously monitors several parameter signals, determining the error in their relationship and initiating corrective action for the error. A typical example is found in the T-average (T_{avg}) controller. The relationship to be maintained for a steady power condition is expressed by the equation

$$T_h + K_1 \frac{dTh}{dt} - K_2 F_S - T_{\text{reference}} = 0$$

where

T_h = reactor outlet temperature
K_1 = adjustable constant in units of time
F_S = steam flow in per cent of full flow
K_2 = adjustable constant for converting steam flow to equivalent degrees Fahrenheit
T_{ref} = adjustable operating reference for T_{avg}

Fig. 8 represents a typical block diagram for this computer.

Normally T_h and T_c (reactor inlet temperature) diverge linearly and proportionally from the T_{avg} for the full range of power. Having established the value of T_h over this range, the T_h parameter channel supplies one control input to the

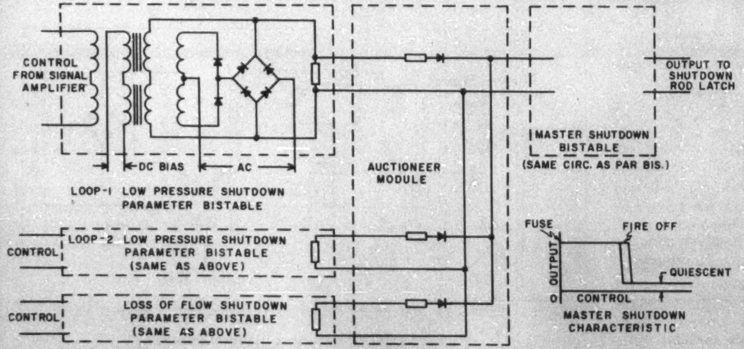
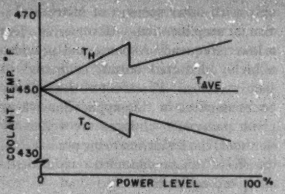

CONTROL FROM SIGNAL AMPLIFIER
DC BIAS
AC
LOOP-1 LOW PRESSURE SHUTDOWN PARAMETER BISTABLE
CONTROL
LOOP-2 LOW PRESSURE SHUTDOWN PARAMETER BISTABLE (SAME AS ABOVE)
CONTROL
LOSS OF FLOW SHUTDOWN PARAMETER BISTABLE (SAME AS ABOVE)
AUCTIONEER MODULE
OUTPUT TO SHUTDOWN ROD LATCH
MASTER SHUTDOWN BISTABLE
(SAME CIRC. AS PAR BIS.)
FUSE
FIRE OFF
OUTPUT
QUIESCENT
0
CONTROL
MASTER SHUTDOWN CHARACTERISTIC

Fig. 7. Typical master shutdown controller channel



RELATIONSHIP OF T_H & T_C TO POWER LEVEL FOR A TWO SPEED PUMP SYSTEM

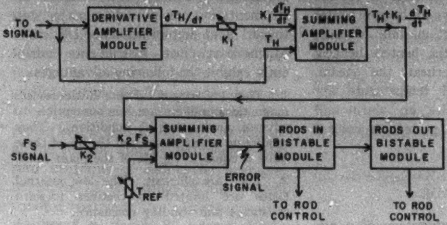


Fig. 8. Typical T_{avg} controller block diagram

computer. An amplifier module extracts the derivative of the T_h signal (effective only during power changes) and supplies this signal to a summing amplifier module where it is combined with the T_h signal to provide the complete temperature expression at any instant for the foregoing equation. The temperature expression is transmitted to a second summing module where it is combined with the reference signal and the signal representative of the steam flow in a subtractive direction to provide the error signal. The error signal is passed on to two bistable controller modules where for a preset positive error the bistable designated RODS IN will initiate a slow insertion of the control rods until the error becomes zero, or for a negative error RODS OUT will be initiated for proper corrective action.

Instrumentating the Reactor Coolant System

From Fig. 1 it will be noted that temperature detectors are located in the main coolant loops to measure T_h and reactor inlet temperature T_c. The detectors are located on the heat exchanger side of the loop isolation valves for plant reliability considerations. In reactor plants of more than two loops, temperature detectors are also located in the steam generator T_{SG}.

The temperature measurements are utilized as follows:

1. The measurements provide a direct measurement of the temperature of the fluid. Wide-range (WR) temperature measurements, from cold start up to plant-operating temperature, are used to provide indication during plant start-up. Narrow-range (NR) temperature measurements made in the normal operating range of the plant are used for indication, and for reactor safety considerations, to provide automatic cutback of reactor power in the event of high T_h.

2. In reactor plants of more than two loops the NR T_h and NR T_c measurements may be auctioneered and the highest T_h and T_c measurements utilized for computing reactor T_{avg} for indication and reactor automatic control. Here also the auctioneered highest T_h measurement may be utilized to provide for automatic cutback of reactor power. Likewise the WR T_c measurements may be auctioneered and compared with individual T_{SG} to prevent accidentally starting up a coolant pump of a shutdown loop and pumping cold water into the reactor. The cold water, serving as a better moderator than the normally hotter water, could cause excessive reactivity sufficient to damage the core.

Pressure measurements are normally made in the pressurizer and in each of the main loops, as shown in Fig. 1.

The pressurizer pressure measurement is utilized for indication, for control of the pressurizer heaters, and for alarm and reactor protection. The loop pressure measurements are utilized as backups for the pressurizer pressure measurement to provide alarms and reactor protection. The loop pressure detectors are also used for indication only when bringing an isolated loop up to pressure and back on the line. With a loop isolated and depressurized for maintenance it is necessary to refill and bleed the loop and then pressurize it to reduce loading on the gate of the loop valve in order to open a valve and reconnect the loop to the reactor system.

Main loop stop valves during normal plant operation are fully open to permit free flow of coolant. A valve in anything other than the fully open position would restrict the coolant flow through the reactor and would result in excessive temperatures within the reactor.

Main loop stop valves are provided, however, to permit deliberate isolation of a loop from the reactor for loop maintenance purposes or simply to stop the flow of coolant through a loop of a multiloop plant in the event that power requirements are such as to not warrant operating the loop. For either of these latter two cases valves or a valve in anything other than the fully closed position would not be serving its intended purpose. Valve position indications, therefore, either fully open or fully closed, are utilized in reactor safety circuits or coolant pump interlock circuits.

In reactor systems utilizing variable speed pump control, coolant flow is usually varied as a function of reactor power in order to conserve the energy required to drive the pumps. In this type of operation excessive coolant flow simply means insufficient utilization of power. However, insufficient coolant flow means more transport time for coolant to flow through the reactor. This being the case, the coolant will become hotter and could conceivably boil into steam, which in turn could result in damage to the reactor. Coolant flow in the individual loops is therefore measured, totalized, and compared with a reactor power level measurement (neutron level), to provide alarm and core protection in the event that an unsafe power to flow ratio exists.

As stated previously, the pressurized water reactor system is a solid water system except for the pressurizer. Since the pressurizer serves a two-fold purpose: 1. of heating water to generate steam for maintaining pressure on the coolant system, and 2. as a surge tank to take care of the expansion and contraction of water in main coolant system due to changes in average coolant temperature, it is necessary for the operator to maintain the level of water in the pressurizer within upper and lower limits in order that the pressurizer may perform properly under normal plant-operating conditions. A measurement of pressurizer liquid level is therefore made to give the operator an indication of level and to provide alarms and heater power cutoff protection.

The pressurizer temperature measurement serves as cross reference on pressurizer pressure measurement and also as a backup for the pressurizer liquid level heater power cutoff protection. The temperature detector is located just above the uppermost heating element in the pressurizer. Should liquid level become lower than the heaters and the level heater power cutoff protection fail to take appropriate action, the heaters heating the steam will superheat the steam. When the measured temperature approaches unsafe limits, an alarm and heater power cutoff protection control action is initiated.

Conclusions

The outstanding advantage of a nuclear power plant is the ability to operate for extended periods without refueling. Essentially this means power for remote areas and locations presently undeveloped for lack of utility power. Providing power facilities to the areas bares the need for plants of low cost, reliability and freedom from maintenance, and economical operation. Hand in hand with these requirements, reliable and accurate instrumentation can provide a major step toward plant simplification and efficiency. Magnetic-amplifier instruments inherently exhibit the following advantages:

1. They are rugged. Being static devices magnetic amplifiers are less susceptible to adverse environmental conditions. They require no maintenance and the life of these units is equal to the life of the plant itself.
2. They are efficient. The low internal losses are limited to rectifier forward resistance and winding resistance.
3. They have stability. The characteristics of operation are predictable and repeatable within close tolerances, the limiting factor being the aging process of rectifiers. Proper rating selection of the rectifiers minimizes this process. Proper design of temperature-compensating circuits minimize effects of ambient temperature changes. These devices have a high signal-to-noise ratio.

The magnetic amplifier compares favorably with other systems of instrumentation for amplification. However, it offers a lower frequency response and provides a highly distorted output compared to some systems. The effect of these drawbacks is offset in this application where plant parameter changes are relatively slow and, the distortion being predictable, its effects can be cancelled out in alignment and calibration.

The utilization of building-block modules to assemble an instrument channel lends itself to modern production. The flexibility of these modules provides an infinite variety of control and computation functions.

It is because of these requirements that the magnetic amplifier has become a new member of the family of instrument systems. This addition is a natural choice in the nuclear power plant field, regardless of the process utilized.

On the Feedback in Magnetic Amplifiers

II—Combined Magnetic and Electric Feedbacks

L. A. FINZI
MEMBER AIEE

J. J. SUOZZI
ASSOCIATE MEMBER AIEE

A PREVIOUS PAPER[1] has presented a steady-state and transient analysis of various types of feedback, magnetic or electric, applied to a self-saturating bridge or center-tap amplifier. This paper is a brief extension of that treatment to certain combinations of a positive magnetic feedback and a negative electric feedback as used by Geyger in "voltage-balance" and "current-balance" potentiometer-type amplifier circuits.[2-4] The essential principle of these feedback combinations is well-known: The positive magnetic feedback reduces the control winding current, while the negative electric feedback insures stability and linearity without modifying the current requirements of the control winding. Two examples are shown in the schematics of Appendixes I and II. (The symbols and terminology are those used consistently in reference 1 and shall not be restated in detail.)

Steady-State Behavior

Voltage-Balance Circuits

In Appendix I, a positive magnetic load-current feedback is associated with a negative series electric load-current feedback. The network analysis is straightforward; the voltage equations of the various loops and the magnetomotive force (mmf) expressions of the cores are written using half-cyclic averages as the variables and relating the preset flux level φ_p to the half-cyclic average mmf F_p of all currents acting on the resetting core by means of an empirical "core-control" curve $\varphi_p=f(F_p)$. Appendix I presents the expressions of the gross voltage gain K_E, of the current gain K_A, of the input resistance R_i, and of the output resistance R_0 and generator gain $\mathcal{K}$ of the Thevenin equivalent of the output circuit. The expressions are given in terms of the voltage gain K_E^* of the "basic" bridge amplifier without feedbacks. Certain approximations suggested by the relative order of magnitude of the parameters in usual designs, are stated in the Appendixes. It should be noted that the current gain K_A depends solely on the magnetic feedback applied, and that the voltage gain K_E can be varied independently without appreciably affecting K_A.

The general expressions simplify if the self-balancing condition $(1-\beta_m K_E^*)=0$ is imposed. In this case the current gain K_A, and thus also the input resistance R_i and the volt-ampere gain $K_{EA}=K_E(K_A)$, become equal to infinity. The further addition of a suitable bias establishes a type of operation in which the input current $I_i=I_c$, is zero over the range of control, while the output remains dictated uniquely by the control voltage E_c with the voltage gain $K_E=1/\beta_c=R_0/r_{sc}$.

(It is recognized that the instantaneous feedback mmf $N_f i_f$, which has the waveform of the output current i_L added to the constant mmf of the bias, may not match the waveform of the instantaneous mmf required by the resetting core. Hence, instantaneous currents i_c flow even when their half-cycle average I_c is zero. Nevertheless, the zeroing of the average control current in itself remains a very interesting feature.)

Paper 58-1232, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958, and re-presented at the Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted July 14, 1958; made available for printing December 4, 1958.

L. A. Finzi is with the Carnegie Institute of Technology, Pittsburgh, Pa., and J. J. Suozzi is with the Bell Telephone Laboratories, Inc., Whippany, N. J.

This work was performed at the Carnegie Institute of Technology under the sponsorship of the Minneapolis-Honeywell Regulator Company, Brown Instruments Division of Philadelphia, Pa. Parts of this paper are abstracted from a doctoral dissertation by J. J. Suozzi. The authors also wish to acknowledge the friendly suggestions given by W. A. Geyger in informal conversations.

Vol.
Pt. 1
1959

Practically a perfect self-balancing may not be obtained over any broad range. The core-control curve is not a straight line (nor, really, is it a sufficiently accurate criterion in this critical question of "zeroing"); also the conducting rectifiers are not the "constant resistors" assumed in the analysis. Hence the theoretical condition $(1-\beta_m K_E^*)=0$ may not be met consistently, and certain values of I_c, positive or negative at various outputs, remain associated with the amplifier control. The band of these values, however, can be quite narrow, and thus the amplifier, in terms of half-cyclic values, can be compared to an almost "ideal voltmeter." It may be noted that the value of the control circuit resistance R_c has no influence on the operation if $(1-\beta_m K_E^*)$ is zero, but actually the practical departures from zero set an upper limit to the admissible value of R_c in the steady state. The general expression of K_E given in Appendix I, can be rewritten as:

$$K_E=\frac{2fN_gK_\varphi N_c}{[R_c+r_{s2}(1-{}^{r_{s2}}/R_g)][1-\beta_m K_E^*]+\dfrac{r_{s2}}{R_g}2fN_gK_\varphi N_c} \qquad (1)$$

where $(1-\beta_m K_E^*)=\left(1-\dfrac{N_f}{R_f}\dfrac{r_{s1}}{R_g}2fN_gK_\varphi\right)$ does not depend on R_c. For large values of R_c with $(1-\beta_m K_E^*)=\delta\neq 0$, the first term in the denominator of equation 1 can intervene to modify significantly the gain K_E. Since δ varies at various outputs, this results in a loss of linearity. In particular, for negative values of δ, the gain K_E may become infinite or negative, i.e., the amplifier is unstable at certain outputs with large R_c. Other limitations on the value of R_c arise in the matter of transient response, as shall be seen.

CURRENT-BALANCE CIRCUITS

In Appendix II, a positive magnetic load-current feedback is associated with a negative shunt electric load-current feedback. Again the steady-state gains and input and output impedances are given in terms of the basic amplifier gain K_E^* (evaluated with a control circuit resistance $R_c=R_K+2r_c$). Since the electric feedback is applied in shunt with the terminals of the control circuit winding, this circuit differs considerably from that of Appendix I. For instance, here, the voltage gain K_E, and the current gain K_A both vary with changes of the electric feedback. The essential feature of this circuit is that the input resistance $R_i=dE_c/dI_i$ is less than the sum of R_K and $2r_c$.

The general results simplify if the further condition $(1-\beta_m K_E^*)=0$ is imposed. In this case, again the half-cyclic control winding current can be zeroed over the range of operation, if a proper bias is added, and thus the voltage at the control winding terminals is also zero in terms of half-cyclic averages, regardless of the value of the control winding resistance, $2r_c$. In other words, the input resistance R_i reduces simply to $R_i=R_K$, and one can say that the winding resistance $2r_c$ is not seen from the control source. Practical departures of $(1-\beta_m K_E^*)$ from zero may be unavoidable over the range, but the band of actual values of I_c and of the winding terminal voltage $E_T=2r_cI_c$, can be quite narrow; thus the amplifier, viewed from the control winding terminals, $(x-x)$ can be compared to an almost "ideal ammeter." (In this case, the general expression given for K_E shows that the actual departures of $(1-\beta_m K_E^*)$ from zero practically set an upper limit on $2r$, and/or a lower limit to R_K, in terms of linearity and steady-state stability.)

Magnetic and electric feedback combinations, essentially similar to those considered in Appendixes I and II, can be obtained using the load voltage, rather than the load current, to provide either feedback or both. The main difference is in the value of output resistance R_0 that results. Generally the effect of the negative electric feedback on the output resistance dominates; hence R_0 is increased above the output resistance of the basic amplifier, if an electric load-current feedback is used, and is decreased by the use of an electric load-voltage feedback.

Transient Behavior

The self-balancing action described for the steady state is disrupted during transients. The matter of transient response is treated in the following with the same methods used in reference 1 and the temporary disturbances that intervene in the transient are related to the amplifier time constant.

VOLTAGE-BALANCE CIRCUIT

The circuit of Appendix I is considered again, first in general terms and then with the additional self-balancing condition $(1-\beta_m K_E^*)=0$. In general terms, assuming that the slope of the core-control curve K_φ, and thus the basic amplifier gain K_E^*, are constant over the range of outputs covered in the transient, and that the bias is supplied from a current source, the difference equations (with n = a positive integer) are:

$$E_{c(n)}=R_cI_{c(n)}+r_{s2}(I_{L(n)}+I_{c(n)})+2fN_c(\varphi_{p(n)}-\varphi_{p(n-1)}) \qquad (2)$$

$$E_{g(n)}=2fN_g(\varphi_s-\varphi_{p(n-1)})+R_gI_{L(n)}-r_{s1}I_{f(n)}+r_{s2}I_{c(n)} \qquad (3)$$

$$r_{s1}(I_{L(n)}-I_{f(n)})=R_fI_{f(n)}+2fN_f(\varphi_{p(n)}-\varphi_{p(n-1)}) \qquad (4)$$

$$N_fI_{f(n)}+N_cI_{c(n)}+{}_{\text{bias}}=\frac{\varphi_{p(n)}-\varphi_K}{K_\varphi} \qquad (5)$$

φ_K is the intercept of the linearized core-control curve with the φ_p axis; $\varphi_{p(n)}$ is the preset flux level at the end of the nth half-cycle while the voltages and currents are half-cyclic averages of the nth half-cycle. For a constant gate voltage E_g, considering only increments from some initial steady state, and assuming that the control voltage E_c is applied at the beginning of a half-cycle and remains constant thereafter, equation 2 becomes:

$$\zeta E_c=\varphi_{p(n)}-\alpha\varphi_{p(n-1)} \qquad (6)$$

where

$$\zeta=\frac{K_\varphi\dfrac{R_c}{N_c}}{1+v+K_{EN}^*(1-q)}$$

and α, q, and v are given in the Appendix I. Equation 6 has the same form found in all cases of the analysis of reference 1 and its solution is

$$\varphi_{p(n)}=\zeta E_c\left(\frac{1-\alpha^n}{1-\alpha}\right) \qquad (7)$$

A continuous "modular" function $y(t)$ can be determined, which for $t=n\pi/\omega$ $(n=0,\ 1,\ 2,\ 3\ \ldots)$ satisfies the condition $y(t)=\varphi_{p(n)}$. This function for $0\leq\alpha<1$ is an exponential characterized by the time constant,

$$\tau_y=\frac{\pi}{\omega}\,\frac{1}{ln\ 1/\alpha}$$

and

$$\varphi_{p(n)}=\frac{\zeta E_c}{1-\alpha}(1-e^{-\pi n/\omega\tau_y}) \qquad (8)$$

Appendix I also presents an expression for the approximate time constant τ_a of the familiar slow-transient approach, in which the discrete values of the half-cyclic average currents and of the preset flux are substituted with continuous time functions. This time constant is adequate for transients requiring many cycles. If the delaying effect of the feedback-loop coupling is neglected (i.e. $N_f^2/R_f\ll N_c^2/R_c$), τ_a simplifies to the familiar form

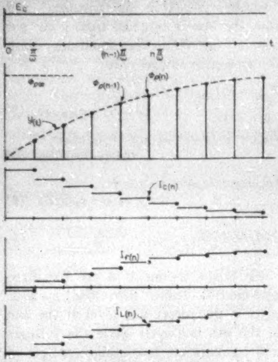


Fig. 1. Transient of preset flux φ_p, control current I_c, feedback current I_f, and load current I_L for a self-balancing amplifier of the voltage-balance type

$$\tau_a=\frac{1}{2f}K_E\frac{N_c}{N_g}$$

in which the voltage bandwidth-gain product is unchanged by the feedback, and the dynamic power gain is given by $D_p=2fK_AN_g/N_c$. The limitations of the approximate approach have been mentioned in section III of reference 1 and shall be re-examined later in this paper.

In perfect self-balancing, the control current $I_{c(n)}$ obtained by solving equations 2, 3, and 4 simultaneously with the condition $(1-\beta_mK_E^*)=0$ is

$$I_{c(n)}=\frac{1}{K_\varphi N_c}\left(1+\frac{N_f^2}{N_c^2}\frac{R_c}{R_f}K_{EN}^*\right)\times (\varphi_{p(n)}-\varphi_{p(n-1)}) \quad (9)$$

while the load current $I_{L(n)}$, and the feedback current $I_{f(n)}$ are respectively

$$I_{L(n)}=-2f\frac{N_f}{R_f}\frac{r_{s1}}{R_g}(\varphi_{p(n)}-\varphi_{p(n-1)})-\frac{r_{s2}}{R_g}I_{c(n)}+\frac{2fN_g}{R_g}\varphi_{p(n-1)} \quad (10)$$

$$I_{f(n)}=\frac{-2fN_f}{(R_f+r_{s1})}(\varphi_{p(n)}-\varphi_{p(n-1)})+\frac{r_{s1}}{(R_f+r_{s1})}I_{L(n)} \quad (11)$$

The simplifications that result for the expressions of α and of τ_a with the self-balancing condition $(1-\beta_mK_E^*)=0$, are seen in Appendix I.

Fig. 1 shows plots of φ_p, I_c, I_L, and I_f in a transient proceeding to higher outputs with a modular time constant $\tau_y\approx 5/2f$ seconds, as an example. The small negative value of $I_{L(n)}$ during the first half-cycle is due to the first two right-hand terms of equation 10. These terms arise from the transient voltage induced during the first half-cycle in the N_f turns associated with the shunt r_{s1}, and to the galvanic coupling of the shunt r_{s2}. Similarly, the negative value of $I_{f(1)}$ is due to the initial predominance of the transient voltage induced in the N_f winding. These effects are not very significant in any practical design. This point is mentioned to show that changes of output, in principle, appear this early during the first half-cycle as due to the magnetic and galvanic couplings.

The values acquired by the control current during the transient are of more serious concern. Considering the exponential rise of $y(t)$, equation 9 indicates that the highest value of $I_{c(n)}$ occurs for $n=1$ and is given by

$$I_{c(1)}=\frac{1}{K_\varphi N_c}\left(1+\frac{N_f^2}{N_c^2}\frac{R_c}{R_f}K_{EN}^*\right)\times \varphi_{p\infty}(1-\epsilon^{-\pi/\omega\tau_y}) \quad (12)$$

Hence, for a given design and thus for a given τ_y, the current $I_{c(1)}$ is proportional to the final steady-state value $\varphi_{p\infty}$, and hereby to the final gross output voltage, $E_\infty=2fN_g\varphi_{p\infty}$.

With a pure d-c signal, the energy demanded from the source in the first half-cycle is

$$U_1=E_c\int_{t=0}^{t=\pi/\omega}i_cdt=\frac{\pi}{\omega}E_cI_{c1} \quad \text{watt-seconds} \quad (13)$$

Since the signal applied is also proportional to the final output, it appears that, for a given design, the energy U_1 (or also the half-cycle average power $\omega/\pi\ U_1=E_cI_{c1}$) is proportional to the square of the swing of output prescribed.

It is interesting also to examine how the transient control currents and powers, and energies associated with a given swing of output, are affected if the time constant τ_y of the amplifier is modified by changes of the amount of negative electric feedback. The influence of τ_y on the control current of the first half-cycle $I_{c(1)}$ is shown in equation 12. For a given $\varphi_{p\infty}$, $I_{c(1)}$ varies proportionally to the factor

$$(1-\epsilon^{-\pi/\omega\tau_y})=\left\{\frac{\pi}{\omega}\frac{1}{\tau_y}-\frac{1}{2!}\left(\frac{\pi}{\omega}\frac{1}{\tau_y}\right)^2+\frac{1}{3!}\left(\frac{\pi}{\omega}\frac{1}{\tau_y}\right)^3-\dots\right\}$$

(and therefore is almost proportional to $1/\tau_y$ for time constants of two half-cycles or more). The dependency of E_c on τ_y is given by:

$$E_c=\frac{1}{K_\varphi N_c}R_c[1+v+K_{EN}^*(1+q)]\times \varphi_{p\infty}(1-\epsilon^{-\pi/\omega\tau_y}) \quad (14)$$

where

$$q=\frac{N_f^2}{N_c^2}\frac{R_c}{R_f}(1+v)$$

and

$$v=\frac{r_{s2}}{R_c}\left(1-\frac{r_{s2}}{R_g}\right)$$

The same factor $(1-\epsilon^{-\pi/(\omega\tau_y)})$ which appears in equation 12 appears here also. However, v is also a variable if changes of τ_y are achieved by changes of negative feedback resulting from changes of the shunt resistance r_{s2}. Decreases in τ_y obtained by increasing r_{s2} cause v to increase (if $r_{s2}<R_g/2$) and therefore E_c increases at a somewhat faster rate than $(1-\epsilon^{-\pi/(\omega\tau_y)})$. Altogether, one can say that the energy of the first half-cycle $U_1=E_cI_{c1}\pi/\omega$, for a given swing of outputs, is approximately proportional to $(1/\tau_y)^2$.

The total electric charge Q drawn from the source during the transient is given by

$$Q=\int_0^\infty i_cdt=\int_0^{\pi/\omega}i_cdt+\int_{\pi/\omega}^{2\pi/\omega}i_cdt\dots+\int_{(n-1)\pi/\omega}^{n\pi/\omega}i_cdt+\dots$$

$$=\sum_{n=1}^{\infty}I_{c(n)}\frac{\pi}{\omega}=\frac{\pi}{\omega}\frac{1}{K_\varphi N_c}\times\left(1+\frac{N_f^2}{N_c^2}\frac{R_c}{R_f}K_{EN}^*\right)\varphi_{p\infty} \quad (15)$$

It is interesting to note that the total charge Q does not vary if τ_y is varied. Hence, the total energy demanded from the source $U_{\text{total}}=E_cQ$ is dependent on τ_y in the same manner as E_c, i.e., U_{total} is approximately proportional to $1/\tau_y$. The results given in the foregoing do not support the suggestion that the response must be rapid in order to minimize transient power drain from the input.[5]

Fast response is readily obtained in the practical design of this self-balancing amplifier. The expression of τ_y indicates that the time constant can be made very short if one is prepared to accept low values of the voltage gain per turn, $K_{EN}=K_EN_c/N_g$. In this respect this amplifier is not different from any one of the amplifiers examined in reference 1. In this self-balancing amplifier, however, a reduction of the voltage gain K_E by an increase in the negative series electric feedback, does not appreciably modify the control-current requirements. In addition, the latter requirements are extremely small to start with because of the positive magnetic feedback. There-

Vol.
Pt. 1
1959

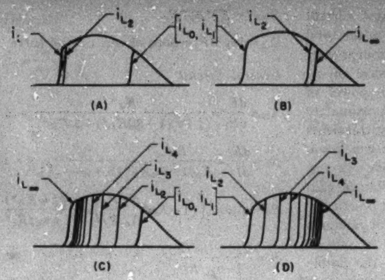


Fig. 2. Transients of load current in a self-balancing Amplifier of the voltage-balance type under the condition $r_{s2}/R_g = N_c/N_g$

A—Up, $R_c = 150$ ohms
B—Down, $R_c = 150$ ohms
C—Up, $R_c = 5{,}000$ ohms
D—Down, $R_c = 5{,}000$ ohms

fore, fast response can be obtained with lower steady-state control power requirements than in other amplifiers.

It may be interesting in particular to consider the speed of response that results for $K_{BN} = 1$. (This condition means $1/\beta_e \; N_c/N_g = 1$, i.e., $r_{s2}/R_g = N_c/N_g$.) In this case the approximate time constant τ_a, neglecting the delaying action of the coupled feedback loop, becomes simply $\tau_a = 1/2f$ seconds. This would imply that the output $E_{(2)}$ of the second half-cycle is only 63.2% of the final steady state. This inference is hardly justifiable since the approximate expression is now strained beyond its rightful limits and the modular expression must be preferred.

Consideration of the modular time constant τ_y, shows that a faster response is obtainable than the approximate expression suggests. For example, for an amplifier with Hypernik V cores $1^1/_2$ by 2 by 1/2 inches, 2-mil tape at 60-cycle and normal voltage, with $K_\varphi \approx 0.15 \times 10^{-3}$ ohms-second, with $N_f = 200$ turns so as to make $\beta_m K_B{}^* = 1$, and with $N_g = 3{,}000$ turns, $N_c = 400$ turns, and $R_g = 1{,}500$ ohms, the condition $r_{s2}/R_g = N_c/N_g$ yields $r_{s2} = 200$ ohms. With $R_c = 150$ ohms, this yields $\tau_y = 0.43$ half-cycle. This means that the preset flux level, $\varphi_{p(1)}$, established at the end of the first half-cycle is about 90% of the final steady state and accordingly the output of the second half-cycle is also about 90% of final. These calculations are in good agreement with the experimental results shown in Figs. 2(A) and 2(B). Fig. 2(A) shows the traces of output currents for a transient starting from an initial output of about 20% of maximum to a final output of 80% of maximum, while Fig. 2(B) applies for the reverse transient between the same limits. The trace of the output current of the second half-cycle indicates $I_{L(2)} \approx 90\%$ of final steady state, while the traces of subsequent half-cycles are no longer distinguishable from the final.

Even faster responses are possible by proper designs and the limiting case of the simple half-cycle delay is obtainable, as has been demonstrated to the authors by Dr. Geyger in his laboratory. However, under the criterion $r_{s2}/R_g = N_c/N_g$, the transient is still an exponential one, even though τ_y may be quite short. In fact, the oscillograms of Figs. 2(C) and 2(D) show that the criterion is not necessarily sufficient to insure a fast response. These oscillograms have been obtained for the same steps of signal voltage as in the cases of Figs. 2(A) and 2(B) respectively, after R_c was increased from 150 to 5,000 ohms. (Self-balancing operation was achieved in this case with only a minor readjustment of the bias current, but with no other changes in the system, and in particular, without changing the condition, $r_{s2}/R_g = N_c/N_g$.) A calculation of the modular time constant τ_y, yields $\tau_y \approx 3$ half-cycles and this agrees rather well with the gradual readjustments of output, half-cycle by half-cycle, that are displayed in the oscillograms. Thus the point is made that values of R_c, which are not too large in terms of steady-state performance, can increase the time constant quite noticeably. The half-cyclic average control current during the first half-cycle of the transient was evaluated as $I_{c(1)} \approx 3$ ma (milliamperes) for the case of Figs. 2(A) and 2(B) while the band of deviation of I_c in steady state was ±50 μa (microampere). For the case of Figs. 2(C) and 2(D), the value of $I_{c(1)}$ was evaluated at about 0.5 ma while the band of deviation of I_c was ± 25 μa in steady state over the range. The transient currents have much larger values than the steady-state bands and this justifies the simplifying assumption of perfect steady-state self-balancing used in this transient analysis.

CURRENT-BALANCE CIRCUIT

The difference equations describing the general system with K_φ constant over the range of outputs, and with a current source bias, are:

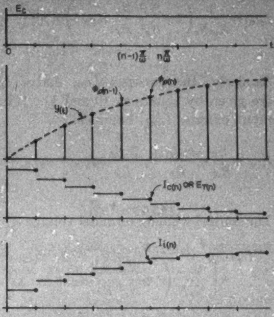


Fig. 3. Transient of preset flux φ_p, control winding current I_c, winding terminal voltage E_T, and input current I_i for a self-balancing amplifier of the current-balance type

$$I_{i(n)} = I_{a(n)} + I_{c(n)} \tag{16}$$

$$E_{c(n)} = R_K I_{i(n)} + 2r_c I_{c(n)} + 2fN_c(\varphi_{p(n)} - \varphi_{p(n-1)}) \tag{17}$$

$$E_{c(n)} = R_K I_{i(n)} + (R_a + r_{s2}) I_{a(n)} - r_{s2} I_{L(n)} \tag{18}$$

$$E_{g(n)} = 2fN_g(\varphi_s - \varphi_{p(n-1)} + R_g I_{L(n)} - r_{s2}(I_{a(n)}) - r_{s1} I_{f(n)} \tag{19}$$

$$r_{s1} I_{L(n)} = (R_f + r_{s1}) I_{f(n)} + 2fN_f(\varphi_{p(n)} - \varphi_{p(n-1)}) \tag{20}$$

$$N_c I_{c(n)} + N_f I_{f(n)} + \text{bias} = \frac{\varphi_{p(n)} - \varphi_K}{K_\varphi} \tag{21}$$

For a constant gate voltage E_g, considering only increments from some initial steady state, and assuming that the control voltage E_c is applied at the beginning of a half-cycle and remains constant thereafter, equation 6 is obtained again,

$$\zeta E_c = \varphi_{p(n)} - \alpha \varphi_{p(n-1)} \tag{6}$$

where

$$\zeta = \frac{(K_\varphi N_c/R_c)\dfrac{R_a + r_{s2}}{R_a + r_{s2} + R_K}}{1 - v + K_{BN}{}^*(1+q)}$$

and q, v, and α are given in Appendix II, under the approximations stated. For $0 \leqslant \alpha < 1$, the solution is again equation 8 where $\tau_y = 1/2f \; 1/ln(1/\alpha)$. Appendix II also presents an expression for τ_a, the time constant of the slow-transient approach.

In perfect self-balancing, the control winding current $I_{c(n)}$ obtained by solving equations 16–21 simultaneously with the condition $(1 - \beta_m K_B{}^*) = 0$, has again the form of equation 9:

$$I_{c(n)}=\frac{1}{K_\varphi N_c}\left(1+\frac{N_f^2}{N_c^2}\frac{R_c}{R_f}K_{EN}^*\right)\times (\varphi_{p(n)}-\varphi_{p(n-1)}) \quad (9)$$

Similarly, the currents $I_{f(n)}$ and $I_{L(n)}$ are given by equations 10 and 11. The input current $I_{i(n)}$, becomes:

$$I_{i(n)}=\frac{E_c}{R_K}-\frac{2r_cI_{c(n)}}{R_K}-\frac{2fN_c}{R_K}(\varphi_{p(n)}-\varphi_{p(n-1)}) \quad (22)$$

The control winding terminal voltage is

$$E_{T(n)}=2r_cI_{c(n)}+2fN_c(\varphi_{p(n)}-\varphi_{p(n-1)}) = \left[\frac{2r_c}{K_\varphi N_c}\left(1+\frac{N_f^2}{N_c^2}\frac{R_c}{R_f}K_{EN}^*\right)+ 2fN_c\right][\varphi_{p(n)}-\varphi_{p(n-1)}] \quad (23)$$

Fig. 3 shows plots of φ_p, I_c, I_i, and E_T for a transient proceeding to higher outputs with a time constant $\tau_y=5/2f$ seconds. $I_{c(n)}$ and $E_{T(n)}$ acquire their highest values in the first half-cycle ($n=1$), and their values $I_{c(1)}$ and $E_{T(1)}$ are obtained by substituting $\varphi_{p\infty}(1-\epsilon^{-\pi/\omega\tau_y})$ for $(\varphi_{p(n)}-\varphi_{p(n-1)})$ in equations 9 and 23. Both quantities are proportional to $\varphi_{p\infty}$, i.e., to the final steady-state output, and to the factor $(1-\epsilon^{-\pi/\omega\tau_y})$ which, for τ_y greater than two half-cycles, can be approximated by $\pi/(\omega\tau_y)$. This shows that smaller time constants, obtained by increasing the negative shunt electric feedback, imply higher initial disturbances of I_c and E_T in the transient.

Fast response again can be obtained if the voltage gain per turn is low[6] (under restrictions which are seen by examining the expression of α given). In this circuit, however, in contrast to the "voltage-balance" circuit, a reduction of K_E by increases of r_{s2} and/or by decreases of R_g, implies a reduction of the current gain K_A at about the same rate.

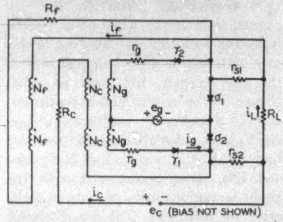
Conclusions

A previous analysis of magnetic amplifiers modified by feedbacks has been extended to two combinations of magnetic and electric feedback with particular reference to self-balancing operation.

For a "voltage-balance" circuit, expressions have been given for the steady-state gains, input and output resistance, and the influence of the control-circuit resistance R_c upon linearity and stability has been mentioned. During transients, self-balancing operation is upset. For a transient, caused by the application of a step of control voltage at the beginning of a half-cycle, it has been shown that the half-cyclic average control current of the first half-cycle as well as the total energy demanded from the source vary approximately as the reciprocal of τ_y, while the half-cyclic average power of the first half-cycle varies approximately as $(1/\tau_y)^2$. The matter of fast transients has been examined with the conclusion that an amount of negative electric feedback which causes the voltage gain per turn to be unity, can result practically in a modular time constant of fractions of a half-cycle and thus in an output response which approximates the simple half-cyclic delay. This is true for low values of R_c; the same voltage gain per turn gives slower transients if R_c is increased.

A similar analysis has been given for a "current-balance" circuit pointing out essential differences that result, in this case, because of the manner in which the negative electric feedback is applied.

Appendix I. Positive Magnetic Load-Current Feedback and Negative Series Electric Load-Current Feedback



Assumed

$$r_{s1}\ll R_f, r_{s1}\ll R_L, r_{s2}\ll R_L, \frac{N_f}{N_c}\frac{r_{s2}}{R_g}\frac{r_{s1}}{R_f}\ll 1$$

Defined

$$\beta_m=\frac{N_f}{R_f}\frac{R_c}{N_c}\frac{r_{s1}}{R_g}, \quad \beta_e=\frac{r_{s2}}{R_g}$$

$$v=\frac{r_{s2}}{R_c}\left(1-\frac{r_{s2}}{R_g}\right), \quad q=\frac{R_cN_f^2}{R_fN_c^2}(1+v)$$

$R_g=R_L+r_{gr}+r_{s1}+r_{s2}$

$R_c=R_K+2r_c$

$r_{gr}=r_g+$rectifier forward resistances

Capital letters, for currents and voltages, indicate half-cyclic averages

$E=E_g-2fN_g(\varphi_s-\varphi_p)=$"gross" output voltage

$E_g=$supply voltage

$\omega=2\pi f=$angular frequency of supply

$K_E^*=$gross voltage gain of bridge circuit without feedback$=2fN_gK_\varphi N_c/R_c$

$K_{EN}^*=K_E^*N_c/N_g$

$K_\varphi=d\varphi_p/dF_p=$slope of core-control curve

General Circuit

$$K_E=\frac{dE}{dE_c}=\frac{K_E^*}{(1+v)(1-\beta_mK_E^*)+\beta_eK_E^*}$$

$$R_i=\frac{dE_c}{dI_c}=\frac{R_cK_E^*}{K_E(1-\beta_mK_E^*)}$$

$$R_0=r_{gr}+r_{s1}\left((1-K_E^*\frac{N_f}{N_c}\frac{R_c}{R_f}\right)+\frac{r_{s2}(1+K_E^*)}{1+r_{s2}/R_c}$$

$$\mathcal{K}=\frac{K_E^*-r_{s2}/R_c}{1+r_{s2}/R_c}$$

$$K_A=\frac{dI_L}{dI_c}=\frac{R_c}{R_g}\frac{K_E^*}{(1-\beta_mK_E^*)}$$

$$\tau_a=\frac{1}{2f}K_E\frac{N_c}{N_g}(1+q)$$

$$\alpha=\frac{K_{EN}^*(1+q)+[\beta_m(1+v)-\beta_e]K_E^*}{1+v+K_{EN}^*(1+q)}$$

$$\tau_y=\frac{1}{2f}\frac{1}{\ln(1/\alpha)}$$

Self-Balancing $(1-\beta_mK_E^*)=0$

$K_E=R_g/r_{s2}$

$R_i=\infty$

$R_0=\mathcal{K}r_{s2}-R_L$

$$\mathcal{K}=\frac{K_E^*-r_{s2}/R_c}{1+r_{s2}/R_c}$$

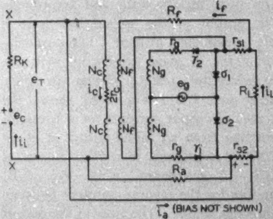
$K_A=\infty$

$$\tau_a=\frac{1}{2f}\left[\frac{R_g}{r_{s2}}\frac{N_c}{N_g}(1+q)\right]$$

$$\alpha=\frac{1+v+K_{EN}^*(1+q)-\beta_eK_E^*}{1+v+K_{EN}^*(1+q)}$$

$$\tau_y=\frac{1}{2f}\frac{1}{\ln(1/\alpha)}$$

Appendix II. Positive Magnetic Load-Current Feedback and Negative Shunt Electric Load-Current Feedback



Assumed

$$r_{s1}\ll R_f, r_{s2}\ll R_L, r_{s1}\ll R_L, \frac{N_f}{N_c}\frac{r_{s2}}{R_g}\frac{r_{s1}}{R_f}\ll 1$$

Defined

$$\beta_m=\frac{R_e\,N_f r_{s1}}{R_f\,N_c\,R_g}$$

$$\beta_e=\frac{R_K}{(R_a+r_{s2}+R_K)}\,\frac{r_{s2}}{R_g}$$

$$v=\frac{R_K^2}{R_c(R_a+r_{s2}+R_K)}$$

$$q=\frac{N_f^2\,R_c}{N_c^2\,R_f}\,(1-v)$$

$$R_g=R_L+r_{gr}+r_{s1}+r_{s2}$$

$$R_c=R_K+2r_c$$

$r_{gr}=r_g+$rectifier forward resistances

General Circuit

$$K_E=\frac{dE}{dE_c}=\frac{K_E^*\left(\dfrac{R_a+r_{s2}}{R_a+r_{s2}+R_K}\right)}{(1-v)(1-\beta_m K_E^*)+\beta_e K_E^*}$$

$$[R_c(R_a+r_{s2})+2r_cR_K]\times$$

$$R_t=\frac{dE_c}{dI_t}=\frac{[1-\beta_m K_E^*]+R_cR_K\dfrac{r_{s2}}{R_g}\;K_E^*}{[(R_a+r_{s2})+2r_c][1-\beta_m K_E^*]+R_c\dfrac{r_{s2}}{R_g}\;K_E^*}$$

$$R_0=r_{gr}+r_{s1}\left(1-K_E^*\frac{N_f}{R_f}\frac{R_e}{N_c}\right)+r_{s2}\left[1+\frac{R_K}{(R_a+r_{s2}+R_K)}\;\frac{K_E^*}{(1-v)}\right]$$

$$\mathcal{K}=K_E^*\left(1-\frac{R_K}{R_a+r_{s2}+R_K}\right)/(1-v)$$

$$K_A=\frac{dI_L}{dI_t}=\frac{\mathcal{K}R_t}{R_L+R_0}=\frac{R_t}{R_g}\;K_E$$

$$\tau_a=\frac{1}{2f}\,K_E\,\frac{N_c}{N_g}\left(\frac{R_a+r_{s2}+R_K}{R_a+r_{s2}}\right)(1+q)$$

$$\alpha=\frac{K_{EN}^*(1+q)+[\beta_m(1-v)-\beta_e]K_E^*}{1-v+K_{EN}^*(1+q)}$$

$$\tau_y=\frac{1}{2f}\;\frac{1}{ln(1/\alpha)}$$

Self-Balancing, $(1-\beta_m K_E^*)=0$

$$K_E=\frac{R_g(R_a+r_{s2})}{r_{s2}R_K}$$

$$R_t=R_K$$

$$R_0=\frac{r_{s2}R_KK_E^*}{(1-v)(R_a+r_{s2}+R_K)}-R_L$$

$\mathcal{K}=$same as Appendix I

$$K_A=\frac{R_a+r_{s2}}{r_{s2}}$$

$$\tau_a=\frac{1}{2f}\left[\frac{R_g(R_a+r_{s2}+R_K)N_c}{r_{s2}R_KN_g}\,(1+q)\right]$$

$$\alpha=\frac{K_{EN}^*(1+q)+1-v-\beta_eK_E^*}{K_{EN}^*(1+q)+1-v}$$

$\tau_y=$same as Appendix I

References

1. On Feedback in Magnetic Amplifiers. Part I—Single Feedbacks, L. A. Finzi, J. J. Suozzi. *AIEE Transactions*, vol. 77, pt. I, 1958 (Jan. 1959 section), pp. 1019–31.

2. Magnetic Amplifiers of the Self-Balancing Potentiometer Type, W. A. Geyger. *Ibid.*, vol. 71, pt. I, Jan. 1952, pp. 383–95.

3. Sixty-Cycle Low-Level Self-Balancing Magnetic Amplifier of the Differential-Feedback Type, W. A. Geyger. "Conference on Magnetic Amplifiers," *AIEE Special Publication T-98*, Aug. 1957, pp. 187–204.

4. The Effective Feedback Factor of Self-Balancing Magnetic Amplifiers, W. A. Geyger. *AIEE Transactions*, vol. 77, pt. I, Mar. 1958, pp. 66–69.

5. The Operation of the Self-Balancing Magnetic Amplifier, A. D. Krall, E. T. Hooper. *Ibid.*, vol. 75, pt. I, Mar. 1956, pp. 79–84.

6. The Operation of the Current-Type Self-Balancing Magnetic Amplifier, A. D. Krall. *Ibid.*, vol. 76, pt. I, Sept. 1957, pp. 380–84.

A Nonsynchronous System for Mobile Record Communications

C. H. STEWART II
MEMBER AIEE

ONE OF THE MOST significant advances in the record communications industry occurred during World War II. This was not so much a technological advance, although this certainly occurred, but a change in operational requirements. The sheer bulk of communications which is required to control adequately and to support a highly mobile combat force plus the difficult communications problems which are posed by such mobile operations early dictated the requirement for more efficient and reliable communications than could be provided by conventional voice circuits. Voice communications are beset by the disadvantages of a relatively low transmission rate and an inherently high transcription error, plus the requirement for a relatively good communications circuit. Consequently, the volume of traffic which is assigned to ripnted record communications has steadily increased until currently the use of voice communications is largely limited to short-range command functions; it is conceivable that, in the not too distant future, even this function will be supplanted by highly portable printing telegraph systems.

The growth of civil printed record communications has developed by an analogous process, although this growth has not been quite as spectacular. The increasing acceptance of data-processing and control systems requires an increasing use of printed communications techniques, to such an extent that the mass of high-speed printed communications which present-day users believe to be absolutely necessary would once have been thought of as a luxury. Many manufacturing processes involve significant field operations which include a sizable mobile force, so that two-way transfer of data is becoming increasingly important to this class of user. The current philosophy of many companies is that this two-way transfer of data naturally extends to the field operations, and it would appear that an ever increasing number of organizations are coming to this point of view. The commercial air carriers are a particularly good example of this philosophy. Presently, their communications requirements are being filled by relatively inefficient voice circuits; however, recent aircraft developments, notably the advent of jet aircraft, have convinced operations personnel that a more efficient means of transferring data is an absolute requirement. This will apparently be handled as a two-stage process, initially by means of a conventional printed

---

Paper 59-8, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted June 26, 1958; made available for printing October 16, 1958.

C. H. Stewart II is with the Bell & Gossett Company, Morton Grove, Ill.

In addition to many people within the Bell & Gossett Company who made significant contributions to the system development, special recognition should be accorded to J. D. Durkee for the basic system concept and for his many continuing contributions. Special credit is also due the consulting firm of Jansky & Bailey, Inc., of Washington, D. C., for their invaluable assistance in the basic explorations and engineering model development.

record communications operation which can later be expanded to a high-speed system for automatic data processing.

The effectiveness of synchronous systems for handling data at high transmission rates with reasonable circuit efficiencies has been demonstrated over a long period. Previously, attempts to extend this technique to services other than land line or fixed plant point-to-point communications circuits have experienced severe difficulties. One of the principal long-term difficulties has been the maintenance of synchronism over the relatively poor communications circuits which generally exist for mobile services. In recent years, several promising approaches have been made toward solving this synchronism problem. One partial solution has been provided through the use of electronic commutation to obviate the requirement of providing a closely controlled rotational rate for the printing telegraph motor. Other approaches add to this the use of a highly redundant binary code and more sophisticated synchronizing means. A radically different approach is that which depends upon a completely nonsynchronous transmission system. Such a system has been developed by the Bell & Gossett Company in co-operation with the Dualex Corporation; and is currently produced under the trade name of the Unitone System. It is this system which will be discussed.

The basic design objective was the provision of a highly reliable mobile extension to a land line or point-to-point record communications–data-handling system. To be completely useful, such a system must provide reliable communications for a circuit which would be marginal or completely unusable for other communications services. As a secondary requirement, it should be possible to implement the system without extensively modifying the existing communications circuit, since a large portion of the potential users currently have some type of communications system. A brief summary of the desired system characteristics follows:

1. Since the system would be essentially an extension to a printing telegraph network, the basic coding scheme should be compatible with the conventional 5-bit printing telegraph code.

2. Circuit reliability requires that the system should employ a slowed-down bit rate for a given transmission rate. This provides an increased sampling period for each bit with a corresponding increase in reliability. This sampling should be provided in a nonsynchronous manner so that the most effective use could be made of the circuit quality which exists at any time.

3. The existing mobile communications circuits are predominately voice circuits; therefore, some type of tone-coded signals would appear to require no equipment modification.

4. Since the system is primarily for mobile applications, a duplex terminal should be provided with low volume, weight, and power consumption.

5. The initial system transmission capacity should represent a reasonable compromise between the requirement for a high-speed data-handling system and the transmission rate which would be required for printed record communications purposes per se. A maximum system capability somewhat in excess of 100 words per minute appeared to be a reasonable compromise, since it would not require a prohibitive development period.

7. The system should be designed with a minimum of complexity, consistent with the desired performance, in order to minimize maintenance problems.

Design Approach

One of the first parameters requiring development was that of the specific coding scheme to be employed. The system requirements were based on a 5-bit code, and the coding system obviously had to be tailored to this requirement if some auxiliary code conversion process were to be obviated. The requirement of a type of tone-coding applied to the 5-bit code would indicate, as an initial approach, the simultaneous transmission of combinations of five audio-frequency tones representing the marking condition of each of the five code elements which comprise the printing telegraph code. Although, at first blush, this would appear to provide "the solution," it is apparent that the advantage of an increased sampling period which thus accrues is essentially nullified by the effective "per-frequency" power reduction resulting from simultaneous transmission. It was also apparent early in the development program that it would be impractical to attempt to utilize the entire character period for sampling since this would require a complex intermediate buffer storage. Thus, if the ultimate performance were to be provided with a minimum of circuit complexity, a different technique would be required. However, it should be noted that a number of prototype units of this so-called Pentone system were constructed which proved very successful for circuits which would normally be considered as marginal for voice communications service.

An evaluation of the relative importance of sampling period versus per-frequency power indicated that the optimum coding system should use alternate transmission of two equal length audio-frequency tones. This would provide maximum effective signal power per tone and also a sampling period approximately four times as long as that of a 5-bit start-stop asynchronous system operating at the same over-all transmission rate.

Fig. 1 illustrates the coding and sampling relations between these two types of systems. The first and second phase tones indicated in Fig. 1 are the bit frequencies from the two families (four and eight respectively) which are transmitted in a consistent time relation, i.e., one of the frequencies of a first family of four always precedes one of a second family of eight. The correspondence between this coding and the 5-bit printing telegraph code occurs in the use of a given frequency to indicate the code element combinations. For example, the four first-phase frequencies represent the combinations of code elements 1 and 2 (1-0, 0-2, 1-2, and 0-0). The second-phase family similarly corresponds to the combinations or code elements 3, 4, and 5. Thus it may be seen that only two bits are used to convey the information contained in the 5-bit code resulting in a significant reduction in effective bit rate. At the receiving terminal, in the stand-by condition, the entire first-phase family of frequencies is exposed to the communications circuit. Statistically, the most reliable operation would appear to be provided by the use of a single first-phase conditioning frequency, at the expense, however, of requiring a second-phase family comprised of 32 frequencies. The increase in reliability was not believed to

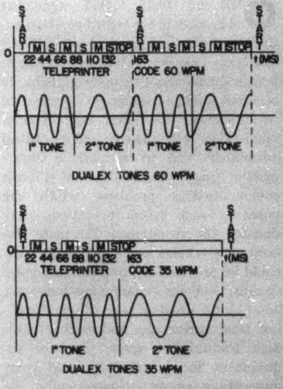


Fig. 1. Comparison of Dualex tone signal with 5-bit start-stop d-c signal

be sufficient to justify this large family of frequencies. The optimum compromise between reliability and frequency complement was found to be the use of a first phase family of four frequencies and a second-phase family of eight frequencies.

To derive the maximum performance which can be provided by the nonsynchronous nature of the system, the receiving terminal should employ a variable bandwidth technique, for, as the circuit deteriorates, it should be possible to reduce the transmission rate and correspondingly automatically reduce the bandwidth of the frequency detectors at the receiving terminal, thus allowing full advantage to be taken of the longer sampling period. This operational characteristic can be closely approximated through the use of conventional filters and rather complex circuitry. The complete realization of this approach requires that very tight tolerances be maintained on the filter characteristics and various threshold levels. This is difficult to achieve in practice. In view of the complexity of this approach, it appeared to be incompatible with the design criteria. A study of the potentialities of resonant reed relays showed that they inherently possess many of the characteristics which are required in the provision of a variable bandwidth system of minimum circuit complexity. This comes about largely from the fact that the output of these relays is a sequence of low-impedance contacts which occur at well defined threshold levels. The absolute value of this threshold is directly dependent upon the method of applying the driving signal. Also, in direct analogy to conventional filters, the response time is inversely proportional to the bandwidth. Thus, it is readily appreciated that the principal ingredients for a variable bandwidth system are inherent to the intrinsic device parameters. One of the characteristics which make the reed attractive, namely, the very high effective Q, makes it impossible to realize keying rates in excess of a few words per minute. The bandwidth–build-up characteristics can readily provide acceptable bandwidths for rise times which are adequate for transmission rates well in excess of 100 words per minute, but the mechanical Q of the reeds is sufficiently high that the reed may never be completely at rest at keying rates over 50 words per minute. The net result is a very erratic build-up time. Extensive testing of the commercially available resonant reed relays showed that none of the units could be successfully utilized for the proposed application.

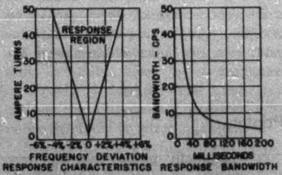


Fig. 2. Resonant reed relay characteristics

The many attractive characteristics of this type of unit appeared to be sufficiently valuable to warrant a development program. It will not be possible in the allotted space to discuss this developmental program; suffice it to say that a practical unit has evolved. Fig. 2 indicates a typical response curve and bandwidth–build-up time relation for the relays as they are currently produced. The solution to the reduction of the reed decay time is currently provided by an electromechanical friction-quenching system. The use of an inertialess measuring technique shows typical damping periods of 10 milliseconds or less. Although this damping factor is more than adequate for sampling rates in excess of 100 words per minute, investigations are continuing toward developing better damping methods.

An examination of the possible methods for driving these units revealed that a constant-current technique would be required in order to achieve the requisite stability and performance. This is because the over-all driven circuit (reed and drive coil) is the responsive in a dual resonant mode about the physical resonance of the reed, so that the impedance of the reed drive coil undergoes a significant fluctuation as the generator frequency is varied through the resonant frequency of the reed. The over-all reed performance is improved, for certain conditions, when the coil circuit is electrically resonated. Because of the intimate relationship between the physical resonance of the reed and that of the reed drive coil, caution must be exercised in establishing the tuning criteria. It is interesting to note that the over-all tuning procedure, which has been utilized for the reed relays for this service, constrains the response characteristic to one which is essentially symmetrical about the resonant frequency rather than the geometric response which would be anticipated for this type of resonant circuit.

Determination of the range of transmission rates to be accommodated by the system immediately fixed the limits for the reed threshold level and, correspondingly, the effective individual bandwidth. The design criteria for accommodation of transmission rates, from approximately 5 to 100 words per minute, indicated that reed threshold response periods ranging from approximately 700 to 25 milliseconds would be required. The resulting maximum bandwidth of approximately 25 cycles allowed for a minimum frequency separation of 40 to 50 cycles, provided that a definite upper limit was placed upon the amplitude of the drive level. The adjustment procedure utilized for the relays results in a threshold-bandwidth-drive level relationship which is essentially independent of the center frequency. As a result of the essentially constant bandwidth, a constant frequency separation of approximately 43 cycles was chosen in lieu of the usual logarithmic separation. The adjustment procedure which yields this constant bandwidth characteristic does result in a variation in absolute sensitivity, i.e., the nose of the response characteristic, which is inversely proportional to the center frequency. This variation in absolute sensitivity becomes the factor which limits the operation under extremely degraded circuit conditions so that it is not feasible to decrease the transmission rate below approximately five words per minute.

The remainder of the development effort was concentrated on the terminal equipment circuitry and printing system. Since the two programs required radically different approaches, they will be discussed separately.

Terminal Unit

Fig. 3 is a simplified block diagram of a typical duplex terminal. Entry to the transmitting converter is on a parallel basis, and is normally accomplished by contact closure of the five code switches on the keyboard or tapereader. The sixth switch performs a dual function of actuating the timing relay and providing simultaneous output of the five code element switches without requiring individually precise adjustments. The five code switches drive five relays whose contacts are arranged in a double "pine tree" matrix deriving the 32-function code.

These contacts select the output of a pair of oscillators whose frequencies correspond to the function which has been selected by the keyboard or tapereader contact combination. The timing relay sequentially switches the amplifier input between the output of the first-phase and the second-phase matrices so that the unique time sequence is maintained. When the input is generated by a tapereader or, if manually generated by a

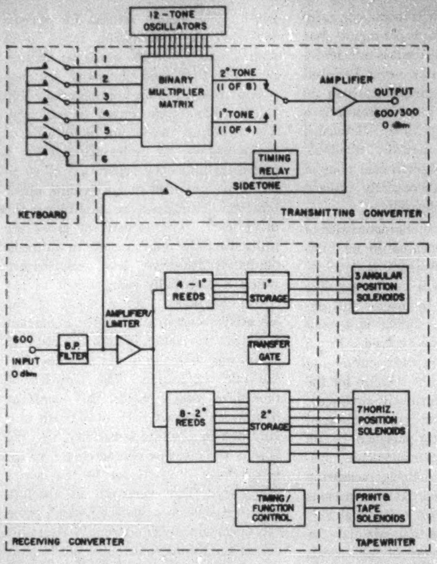


Fig. 3. Simplified block diagram of duplex terminal

trained operator, the time division between the first- and second-phase frequencies is sufficiently accurate that auxiliary timing circuits are not required. When it is anticipated that relatively inexperienced operators will operate the circuit, an auxiliary plug-in timing relay is utilized. This relay limits the length of the first-phase tone and thus minimizes the effect of operator timing error. The amplifier output is squelched during the stand-by period so there is no possibility that tone leak-through will interfere with circuits which are used for voice communications on a shared time basis. A normally open contact is also provided for use with those equipments having a "push-to-talk" mode of operation. This circuit has a release delay of approximately 750 milliseconds, providing for no carrier interruptions for keying rates in excess of approximately five words per minute. Inasmuch as the normal installation would be on existing voice circuits, the nominal output was chosen as zero decibels referred to one milliwatt. Provision is made for monitoring the transmitted copy by means of a side-tone amplifier whose output is well in excess of the input threshold of the receiving converter.

The receiving converter consists of three basic subdivisions: the preliminary filtering and amplification (which provides the drive for the reed detectors); the frequency determining elements (reed detectors); and the storage and timing circuits. It will be noted in Fig. 3 that a limiter-amplifier combination is used to provide the drive to the resonant reed relays. The reason for this is twofold: first, to preclude the possibility of printing erroneous characters, driving power must be limited in order to prevent the reed threshold from being raised to the point where a given reed responds to adjacent frequencies; second, the quasi-capture effect of a broad-band limiter increases the effective signal-to-noise ratio if it is greater than unity and depresses the reed signal power for signal-to-noise ratios less than unity. This limiter effect tends to reduce the drive to the reed detectors and correspondingly the acceptance band for signal-to-noise ratios less than unity so that the effective bandwidth is essentially proportional to the signal-to-noise ratio. Thus, the desired variable bandwidth system performance has been approximated by means of relatively simple circuitry.

The transfer gate initially conditions the first-phase family of reed detectors so that all four are exposed to the communications circuit. Upon recognition of one of the first-phase frequencies, the appropriate code is stored and, simultaneously, the gate conditions the second-phase family of frequencies. Upon recognition of a second-phase frequency, this information is stored and simultaneously the print cycle is initiated. With the present printer, the print cycle consists of approximately 50 milliseconds' positioning time, plus 25 to 30 milliseconds' printing time. The time which is required for the overall sequence of relay operations, plus the print cycle time, requires that it be possible to store succeeding first-phase information during the print cycle in order to achieve transmission rates in excess of approximately 85 words per minute. This is accomplished by utilizing a compound first-phase storage, i.e., buffer storage plus output storage.

It will be noted that a 12-wire output is required by the printer (10-code element plus print and return leads). To accommodate the use of tape perforators or other auxiliary equipment, the output storage incorporates a binary divider so as to provide a 5-wire parallel output. It should also be pointed out that since the storage is accomplished by means of relays, the circuit is arranged so that only two heavy-duty relays make or break the output current, thus reducing contact wear.

As a consequence of the high use factor for the radio-frequency spectrum, the incidence of co-channel and adjacent channel interference is quite high. To provide protection against interfering components' falling within the signal band, an additional timing circuit was incorporated in the transfer gate. The function of this circuit is to clear the first-phase storage if a second-phase frequency is not recognized within a certain period of time. This time is determined by the length of the first-phase frequency; thus, the period which is allowed before clearing is determined by the effective transmission rate. This circuit has greatly reduced the sensitivity of the system to interfering signals although it is manifestly impossible to preclude such errors completely by this approach.

Early in the design program, it was quite evident that a stabilized low-impedance voltage source would be required for the gating and timing functions. This is provided by a regulated supply whose output impedance is less than 5 ohms for a frequency range of direct current to 200,000 cycles, so that no instability results from any pulses which may be encountered.

Printer

It was initially decided that the overall system requirements dictated a fresh approach to the printer problem, rather

Vol.
Pt. 1
195

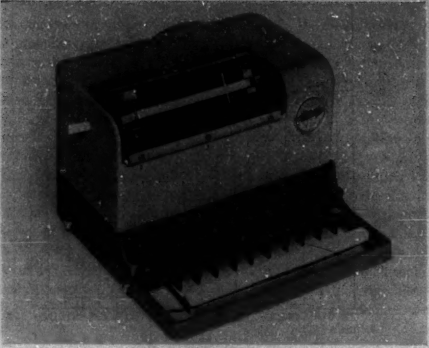


Fig. 4. Tapewriter and keyboard

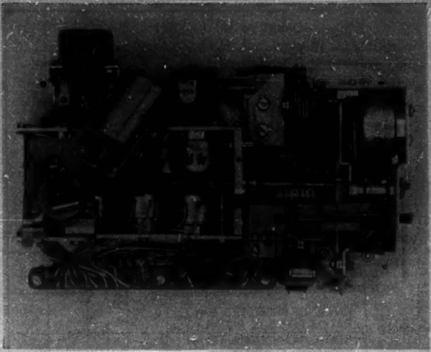


Fig. 5. View of tapewriter with cover removed

than the attempt to modify an existing unit. After thorough exploration of the many possible approaches, it was decided that the requirement for satisfactory operation under difficult environmental conditions, plus the requirement for minimum size, weight, and power consumption could best be met by a tape printer configuration. The basic mechanism consists of rotary solenoids driving a type segment arranged in a 4-by-8 matrix. This approach results in minimum power consumption since no power is required except during the printing operation. Fig. 4 shows the assembled printer and keyboard. The type segment arrangement as a 4-by-8 matrix requires seven active horizontal positions and three active angular positions. The seven rotary solenoids and associated linkage constitute the subassembly in the left foreground of Fig. 5. The angular positioning subassembly, which consists of the three solenoids and associated linkage, comprises the assembly at the extreme right. It will be noted that the type font is reversed so that the tape is between the operator and the type font. This configuration provides maximum visibility. To reduce further the mechanical movements to a minimum, the conventional paper tape plus inked ribbon combination has been discarded in favor of impact-sensitive tape. Upper and lower case selection (figures-letters shift), is accomplished by rotating the tape guide 11 degrees. This subdivides each of the 32 type font positions into either the figure or letter characters.

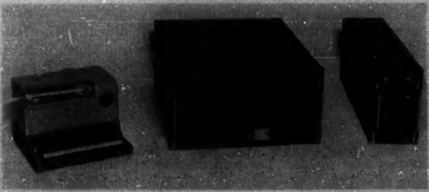
At present, the driving current is applied as a discontinuous pulse; however, a subsequent model of the system will provide tailored driving pulses so that all of the movements will be subjected to simple harmonic acceleration.

Results and Conclusions

Fig. 6 shows a typical duplex terminal developed by the program. The converter and power supply are packaged in the standard ATR (American Television & Radio Company) configuration since it was anticipated that the initial use would be in an air-borne environment due to the stringent communication requirements of this service. The major objective of the design program was the development of a system which approached the performance capability of the basic concept, making it possible to acquire a useful quantity of data from actual field service. A miniaturization program is being initiated which should result in significant improvements in the over-all performance and major reductions in the displacement, weight, and power consumption. The maintenance-free life should also be greatly extended since solid-state switching is planned for all of the circuits.

The over-all system performance has been extensively measured in the laboratory. The results of a typical test are shown as Fig. 7. This test utilized a white noise source, band limited to the signal band by an inductance-capacitance filter with 40 decibels per octave skirt selectivity. The rms noise envelope is plotted versus the maximum transmission rate for an error count equal to or less than 1%. It will be noted that the curve directly shows the quasi-capture effect of the limiter on the reed detector bandwidth and corresponding threshold build-up time. It will also be noted that, because of the maximum speed capability of the system, the deviation from the logarithmic relation between the signal-to-noise ratio and channel capacity increases at the higher transmission rates. The zone which is shown as 50% error for 5-bit synchronous systems is the predicted signal-to-noise ratio for this maximum error rate as described by Masonson[1] and Montgomery.[2] The indicated range includes both amplitude modulation and frequency shift keying for noise, band-limited to the transmission band. Although the predicted performance of the binary systems is based on several simplifying assumptions, the improvement in performance provided by the nonsynchronous approach is believed to be significant.

Fig. 6. Type U2-A typical duplex airborne terminal developed by the program



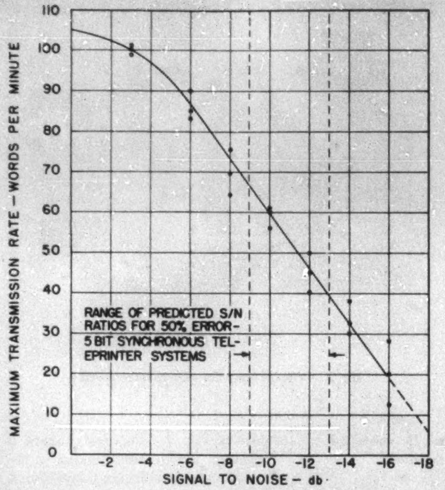


Fig. 7 (left). Maximum transmission rate for 1% error

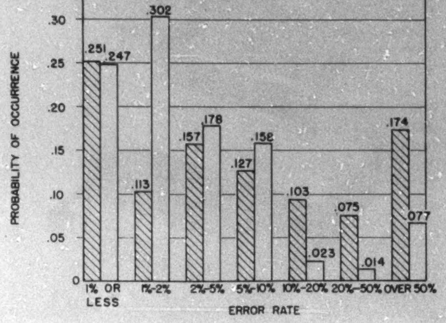


Fig. 8 (above). Wright Air Development Center flight tests, 1956. Shaded areas—frequency shift keying. Unshaded areas—Unitone DSB-AM

The "proof of the pudding" for any system, of course, is the performance under actual circuit conditions. The system has been flight tested by the Department of Defense with various test objectives. Two of the more quantitative test programs will be briefly described.

One of the earlier tests was conducted by the Air Weather Service in February 1956.[3] The purpose of the test was to compare the system performance with that of voice and manual CW (continuous wave) techniques. Inasmuch as it was desired that the program also furnish operational suitability information, the transmissions consisted of the operational traffic which was generated by the weather reconnaissance flights from Kindley Air Force Base, Bermuda. The weather reconnaissance tracks flown were essentially a triangular path of approximately 1,100 miles per leg. Each message was transmitted in sequence as rapidly as possible by manual CW, voice, and the Unitone system in order to minimize variations in propagation conditions. Conventional double side-band amplitude modulation was utilized for the voice and Unitone transmissions with equal carrier power maintained for all modes of transmission. The effective power output from the air-borne transmitter was less than 100 watts and the ground transmitters were operated at a nominal power level of 3,000 watts. The transmissions were scored on a per-character basis by directly comparing the transmitted and received copy. The following conclusions were based on an analysis of the test results. The voice transmissions were sufficiently ineffective that no relative comparisons were attempted. The Unitone transmissions were found to be as effective as the manual CW, and for very degraded circuit conditions appeared to be somewhat superior, especially regarding effective transmission rate. This was indeed gratifying since manual CW has long been the standard of comparison for the performance of degraded circuits.

During June and July of 1956, the Communications and Navigations Laboratory at Wright Air Development Center conducted a more quantitative test, to compare the performance of the Unitone system with that of a frequency-shift keying teleprinter system.[4] Consequently, all transmissions which were utilized for scoring purposes used prepared test tapes and mechanical transmission for both systems. Transmissions, in each mode, were made sequentially in 3- to 5-minute increments in order to minimize variations in propagation conditions. The aircraft radio-frequency equipment was an *ARC-21* transceiver plus an *XA-4* exciter for the frequency shift keying transmissions. The test flights varied from 400 to a maximum of 1,700 miles. The operational frequencies for each period were selected at the three extremes of frequency usage. During each hour, one series of transmissions was made on the optimum traffic frequency, the second on a frequency which was above the predicted maximum usable frequency, and the final sequence utilized a frequency which was below the predicted lowest useful frequency for the particular path. The circuit was often further degraded by adjusting the azimuthal orientation of the ground antenna system away from the propagation path.

A total of 74,362 characters, for each mode of transmission, was retained for scoring purposes. The general conclusions resulting from the test programs were that, for good circuit conditions, the performance provided by both techniques was essentially equivalent; but that, for deteriorated circuit conditions, the Unitone system, in the aggregate, provided a better "per-character" reliability of 15% to 25%. Fig. 8 shows the results of the flight test program, presenting the data as probability of occurrence for various error rates for each mode of transmission. The error rates are grouped logarithmically since this provides the best indication of the over-all system usefulness. As would be expected for error rates of 1% or less, the two systems provide an equivalent probability which indicates simply that, for a reasonably good transmission medium, the choice of technique is not very important; however, for deteriorated conditions the differences are more pronounced. For example, the Unitone system, for the wide range of circuit conditions, exhibited the highest probability for error rates of 1% to 2% with the probability of higher error rates decreasing rapidly. The frequency shift keying system, on the other hand, exhibits a probability peak for a 2% to 5% error rate with a relatively gradual decrease in probability for higher error rates. It is also informative to compare the probability of occurrence for error rates in excess of 50%. Although it would certainly be a grave error to draw sweeping conclusions from the limited test data, it is believed that they are an indication of the potentiality of this technique. It is believed that the differences would have been even greater had the modulation system been other than double side-band amplitude modulation. For example, a series of propagation-modulation tests which were conducted by the Bell & Gossett Company over a 600-mile path indicated that the use of a narrow-band frequency modulation technique such as proposed by Arguimbau resulted in an effective power gain of approximately 6 decibels over double side-band amplitude modulation.

The system described is currently being produced for use by the Air Defense Command as the primary air-ground communications system for the Airborne Early Warning environment. As a part of this program, a duplex serial to parallel adapter has been developed and is in production for installation in the ground environment. The adapter provides direct entry into a conventional start-stop 5-bit telegraph system. Decreased maintenance requirements are provided by the use of modular packaging techniques and solid-state circuitry throughout.

The laboratory testing and, to date, the somewhat limited flight testing of the Unitone nonsynchronous system indicates that this approach provides increased reliability for difficult communications circuits. The equipments which are currently under development should provide even better performance which more closely approaches the ultimate system capability.

References

1. Binary Transmission Through Noise and Fading, M. Masonson. *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. II, 1957, pp. 69–82.
2. A Comparison of Amplitude and Angle Modulation for Narrow-Band Communication of Binary-Coded Messages in Fluctuation Noise, G. F. Montgomery. *Proceedings*, Institute of Radio Engineers, vol. 42, 1954, pp. 447–54.
3. *Report*, Manufacturer's Demonstration, Dualex Unitone Tapewriter. Kindley Air Force Base, Bermuda, Hq. Air Weather Service, Military Air Transport Service, Washington, D. C., May 1956.
4. Flight Test of Bell and Gossett Dualex Teletypewriter Equipment. *Memorandum Report*, Bell & Gossett, New York, N. Y., Sept. 5, 1956.

Industry Schools Its Engineers

C. E. WALDNER
MEMBER AIEE

WHY IS INDUSTRY starting and expanding programs to educate its young engineers beyond the level where formal education left off? Why is it reviewing the education of its older engineers and giving them further training? Are not the great universities offering their students the proper curriculums?

The answer, of course, is that, while the universities are doing a fine job, technologies in almost every field are changing and advancing so rapidly that a formal education begins to be out of date in a period of only a few years. For instance, that which is no more than a theory during the college student's freshman year may become a practical project and go into production during his senior year, taking only 4 years to materialize. Consequently, industry, to keep pace with these advances and changes, and to ensure that there are those available who understand and can handle the new technologies, must take the initiative in training its people, particularly in areas requiring specialized education.

Training in Industry

In the telephone business, one is constantly impressed by the depth and scope of present-day formal education in communications. Many an "oldtimer," who himself may have been able to take only one or two elective hours of communications in his mechanical and electrical engineering college, is struck by the knowledge and the fluency of those who major today in communications, in discussing theories involving transistors, waveguides, radio, etc. Yet, in the telephone business particularly, the need for such educated personnel far exceeds the supply.

During the last two decades, industries that have entered the field of electronics have multiplied many times and their demand for these college graduates has become great. It is not the telephone systems alone that require their services: now, there are data-processing and transmission, guided missiles, special communications for jet planes, not to mention the much advertised trip to the moon.

There is also the need to consider for such training the present Bell System engineering manpower, a core of experienced and capable people who, during 1958, added plant that requires an expenditure of over 2 billion dollars. The experience of these men (and there are some women engineers) ranges from less than 1 to more than 40 years. Some are engineering graduates; others are without college degrees. Among the former, very few have majored in communications, and a relatively small number are conversant in the newest arts. (This does not apply to those whose hobbies are in connection with radio, nor to those who have gained knowledge through self-study and actual experience, nor to the highly knowledgeable experts of the Bell Laboratories.)

Yet it is among these existing engineers, where planning for the future more than ever before is rapidly leading into the new electronic arts, that the need for knowledge and information is most acutely felt. For example, formerly open-wire and cable afforded the only conversational channels; now carrier and radio are taking a larger and larger share. Any circuit that operated efficiently on voice frequency, say, between 200 and 3,000 cycles, used to meet requirements; now kilocycles and megacycles are involved, with

---

Paper 58-1235, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted July 13, 1958; made available for printing January 28, 1959.

C. E. Waldner is with the New York Telephone Company, New York, N. Y.

Special credit is due to Cornell University for accepting and carrying out this program. Acknowledgment is also due to the following for their assistance in organizing the training course: Prof. W. H. Erickson, director, School of Electrical Engineering, and Dr. H. G. Smith, professor of communications, Cornell University; Dr. J. K. Lamond, C. N. Anderson, P. A. Jeanne, and L. C. Roberts, former Bell System engineers, now instructors at Cornell University; F. H. Tallman and M. B. Zaret, professors of communications, Cooper Union; W. R. Silvey, S. Hayward, and R. C. Crawshaw, engineers of outside plant and transmission, and H. M. Wilson, transmission engineer, New York Telephone Company.

transmission making essential circuit refinements that were previously unheard of. And, in a relatively few years, switching systems will no longer exclusively use magnetic relays but rather diodes, transistors, and cold cathode tubes. And there is more to come!

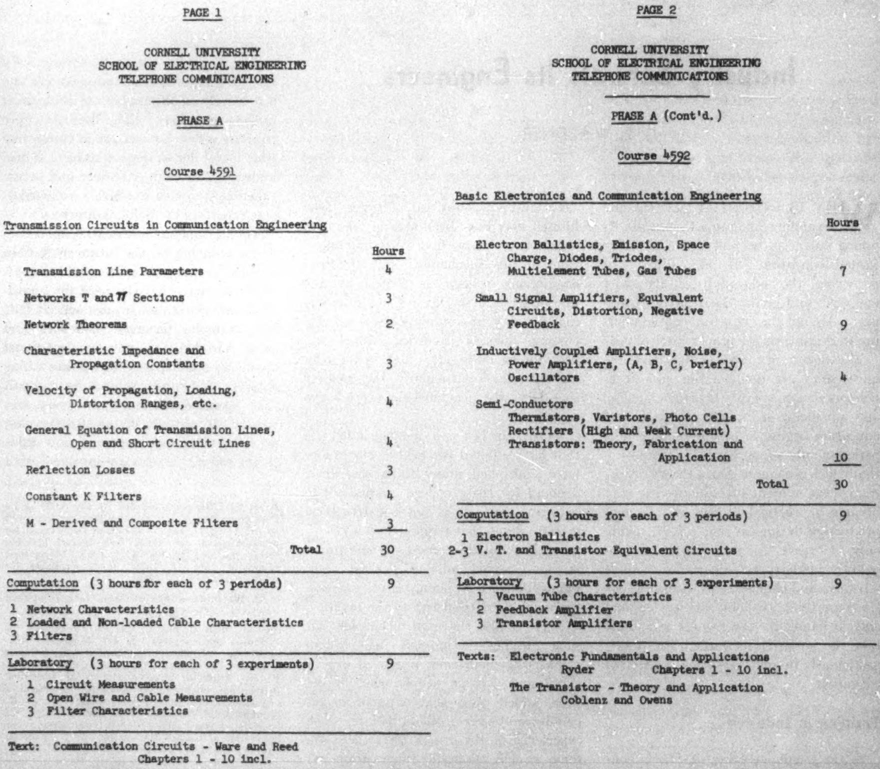
The practice of training engineers has been active among Bell System companies for some time. It has been applied to equipment engineers, who lay out, order, and oversee the installation of switching and auxiliary apparatus; the outside plant engineer, who is responsible for poles, cables, and subways, and to others of the engineering fraternity.

The advent of modern electronics, however, has revolutionized the field of telephony. This has brought the need for a whole new technical orientation which can be met only through a new, specialized educational program, the development of which has required the establishment of new evaluation and training methods and the setting up of additional educational facilities. In telephone parlance, the electronics engineering trainees are known as communications engineers in a field which telephone people, with justification, refer to as the New Arts.

The telephone industry, then, like many other industries, is increasingly taking on the aspect of an educational, as well as a service, institution. Telephony is progressing at its most rapid rate in history. Therefore it is essential that the industry foster the education of its engineers in the new arts of communication. It is only through continuous study of the latest sciences that the communications world of tomorrow can be designed, built, and perfected.

The New York Telephone Company Program

Since early in 1956 the New York Telephone Company has given in-hour, off-the-job training over periods of from 1 to 3 weeks to about 800 of the central office, outside plant, and customer equipment engineers. During the last year it has had 2,300 enrollments in a correspondence school. This type of training will be continued as new equipment is provided and practices change, and also as engineers are transferred or new ones are brought into the business. To meet the needs for engineering in electronics, however, it was felt necessary to obtain the facilities of a large university in the curriculum of which communications played a predominate part. After considerable investigation, a school for communications



PAGE 1

CORNELL UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING
TELEPHONE COMMUNICATIONS

PHASE A

Course 4591

Transmission Circuits in Communication Engineering

	Hours
Transmission Line Parameters	4
Networks T and π Sections	3
Network Theorems	2
Characteristic Impedance and Propagation Constants	3
Velocity of Propagation, Loading, Distortion Ranges, etc.	4
General Equation of Transmission Lines, Open and Short Circuit Lines	4
Reflection Losses	3
Constant K Filters	4
M - Derived and Composite Filters	3
Total	30
Computation (3 hours for each of 3 periods)	9
Laboratory (3 hours for each of 3 experiments)	9

Computation:
1 Network Characteristics
2 Loaded and Non-loaded Cable Characteristics
3 Filters

Laboratory:
1 Circuit Measurements
2 Open Wire and Cable Measurements
3 Filter Characteristics

Text: Communication Circuits - Ware and Reed
Chapters 1 - 10 incl.

PAGE 2

CORNELL UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING
TELEPHONE COMMUNICATIONS

PHASE A (Cont'd.)

Course 4592

Basic Electronics and Communication Engineering

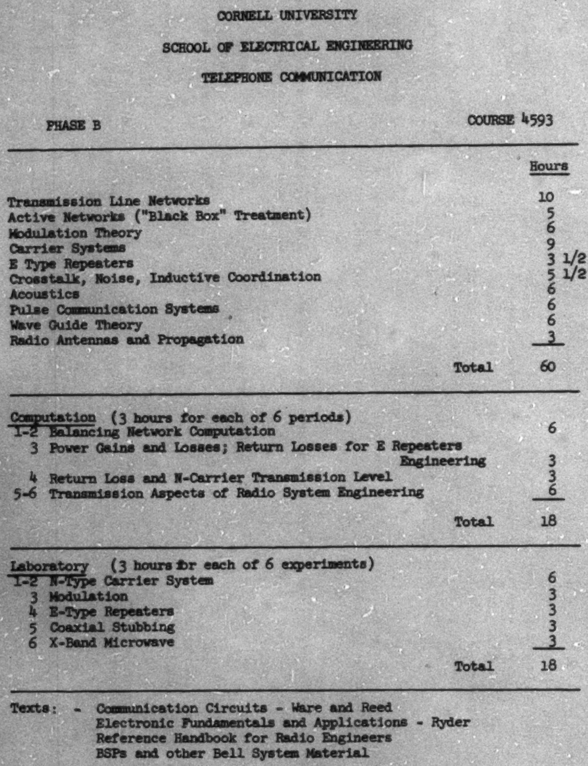
	Hours
Electron Ballistics, Emission, Space Charge, Diodes, Triodes, Multielement Tubes, Gas Tubes	7
Small Signal Amplifiers, Equivalent Circuits, Distortion, Negative Feedback	9
Inductively Coupled Amplifiers, Noise, Power Amplifiers, (A, B, C, briefly) Oscillators	4
Semi-Conductors Thermistors, Varistors, Photo Cells Rectifiers (High and Weak Current) Transistors: Theory, Fabrication and Application	10
Total	30
Computation (3 hours for each of 3 periods)	9
Laboratory (3 hours for each of 3 experiments)	9

Computation:
1 Electron Ballistics
2-3 V. T. and Transistor Equivalent Circuits

Laboratory:
1 Vacuum Tube Characteristics
2 Feedback Amplifier
3 Transistor Amplifiers

Texts: Electronic Fundamentals and Applications
Ryder Chapters 1 - 10 incl.
The Transistor - Theory and Application
Coblenz and Owens

Fig. 1. Review of general principles of transmission lines and basic electronics; phase A of NYTEL (New York Telephone Company) course



CORNELL UNIVERSITY

SCHOOL OF ELECTRICAL ENGINEERING

TELEPHONE COMMUNICATION

PHASE B — COURSE 4593

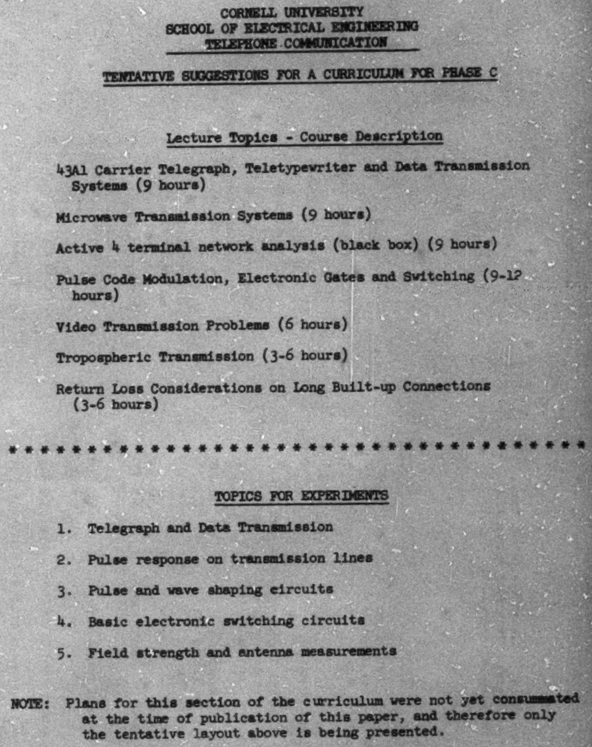
	Hours
Transmission Line Networks	10
Active Networks ("Black Box" Treatment)	5
Modulation Theory	6
Carrier Systems	9
E Type Repeaters	3 1/2
Crosstalk, Noise, Inductive Coordination	5 1/2
Acoustics	6
Pulse Communication Systems	6
Wave Guide Theory	6
Radio Antennas and Propagation	3
Total	60

Computation (3 hours for each of 6 periods)	
1-2 Balancing Network Computation	6
3 Power Gains and Losses; Return Losses for E Repeaters Engineering	3
4 Return Loss and N-Carrier Transmission Level	3
5-6 Transmission Aspects of Radio System Engineering	6
Total	18

Laboratory (3 hours for each of 6 experiments)	
1-2 N-Type Carrier System	6
3 Modulation	3
4 E-Type Repeaters	3
5 Coaxial Stubbing	3
6 X-Band Microwave	3
Total	18

Texts: - Communication Circuits - Ware and Reed
Electronic Fundamentals and Applications - Ryder
Reference Handbook for Radio Engineers
BSPs and other Bell System Material

Fig. 2. Phase B of NYTEL course



CORNELL UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING
TELEPHONE COMMUNICATION

TENTATIVE SUGGESTIONS FOR A CURRICULUM FOR PHASE C

Lecture Topics - Course Description

43A1 Carrier Telegraph, Teletypewriter and Data Transmission Systems (9 hours)

Microwave Transmission Systems (9 hours)

Active 4 terminal network analysis (black box) (9 hours)

Pulse Code Modulation, Electronic Gates and Switching (9-12 hours)

Video Transmission Problems (6 hours)

Tropospheric Transmission (3-6 hours)

Return Loss Considerations on Long Built-up Connections (3-6 hours)

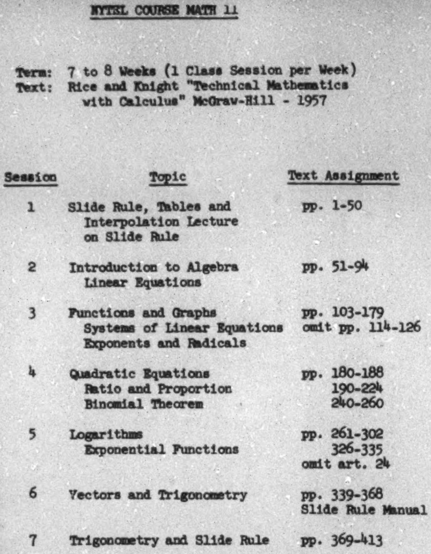
* *

TOPICS FOR EXPERIMENTS

1. Telegraph and Data Transmission
2. Pulse response on transmission lines
3. Pulse and wave shaping circuits
4. Basic electronic switching circuits
5. Field strength and antenna measurements

NOTE: Plans for this section of the curriculum were not yet consummated at the time of publication of this paper, and therefore only the tentative layout above is being presented.

Fig. 3. Phase C of NYTEL course

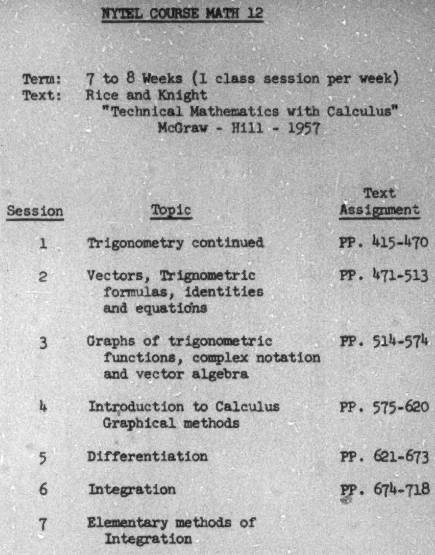


NYTEL COURSE MATH 11

Term: 7 to 8 Weeks (1 Class Session per Week)
Text: Rice and Knight "Technical Mathematics with Calculus" McGraw-Hill - 1957

Session	Topic	Text Assignment
1	Slide Rule, Tables and Interpolation Lecture on Slide Rule	pp. 1-50
2	Introduction to Algebra Linear Equations	pp. 51-94
3	Functions and Graphs Systems of Linear Equations Exponents and Radicals	pp. 103-179 omit pp. 114-126
4	Quadratic Equations Ratio and Proportion Binomial Theorem	pp. 180-188 190-224 240-260
5	Logarithms Exponential Functions	pp. 261-302 326-335 omit art. 24
6	Vectors and Trigonometry	pp. 339-368 Slide Rule Manual
7	Trigonometry and Slide Rule	pp. 369-413

Fig. 4. Mathematics 11 of NYTEL course



NYTEL COURSE MATH 12

Term: 7 to 8 Weeks (1 class session per week)
Text: Rice and Knight "Technical Mathematics with Calculus" McGraw - Hill - 1957

Session	Topic	Text Assignment
1	Trigonometry continued	PP. 415-470
2	Vectors, Trignometric formulas, identities and equations	PP. 471-513
3	Graphs of trigonometric functions, complex notation and vector algebra	PP. 514-574
4	Introduction to Calculus Graphical methods	PP. 575-620
5	Differentiation	PP. 621-673
6	Integration	PP. 674-718
7	Elementary methods of Integration	

Fig. 5. Mathematics 12 of NYTEL course

engineers was eatablished at Cornell University, starting with the school year beginning in September 1957. Each student attends three 3-week sessions spaced about 4 to 6 months apart, with a maximum of 24 students in his class. The sessions have been planned to run continuously for 2 years, excluding summer months and several regular college recesses, and to terminate in June 1959. Altogether, 200 engineers are receiving training.

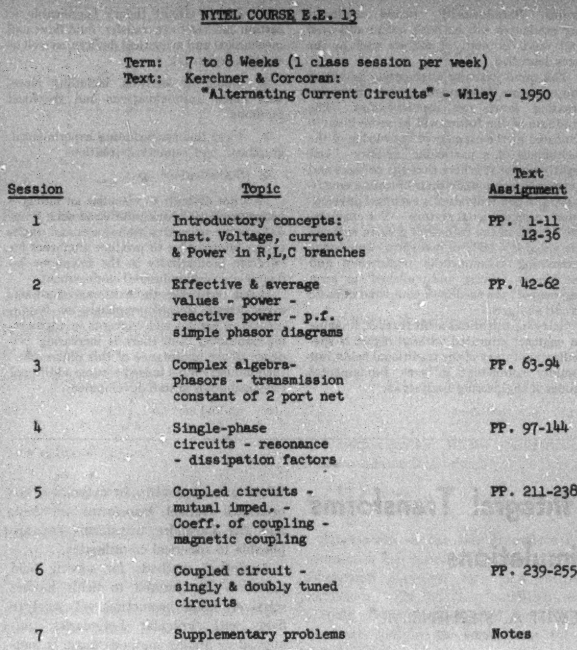
Many problems were encountered in organizing the training course. It was necessary to bring together into one class men from widely different eras in college education, with varying degrees and types of education and of company experience. Instructors with considerable knowledge of these subjects were required. These and other requirements had to be met before a successfully operated school could be realized.

It was recognized that for successful training, any great scholastic level discrepancies among the students would have to be avoided. For instance, if higher mathematics were the subject being taught, a man with little education would retard the progress of the entire class. Therefore, to measure qualifications for this communications training program, a 4-hour examination was developed and given to over 300 engineers, including also some Plant Department supervisory people. It consisted of two parts, mathematics and electronics, the former being a prerequisite, of course, in any studies of the latter. After a large sample of the answer papers had been reviewed in order to determine some representative and practical grouping, all papers were marked and the engineers segregated into three categories: 1. men not qualified to enter into the high standard of education contemplated until they had taken prerequisite courses in mathematics and in a-c circuits, 2. men qualified after taking the course in a-c circuits, and 3. men qualified without the advanced training recommended for categories 1 and 2.

Another very important function of this examination was to determine the curriculum, and the depth of training needed. There would have been little point for the average student in going into higher mathematics or the most recently advanced theories of electronics if he actually required training to bring him to a standard below this level. Also, the college graduates of years ago, without daily usage of their undergraduate learning, needed refresher courses. The company considered these factors when selecting the subject matter for instruction. Therefore, for the initial period, or phase A of training, a review of the general principles of transmission lines and basic electronics was concentrated upon. The plan was to cover this in 3 weeks with 60 hours of classroom work and 18 hours each of computation and laboratory periods; see Fig. 1.

Other subjects were contemplated at this time but it was decided not to set up a further definitive program until later. It was felt that it would be helpful, in laying out the studies subsequent to phase A, if first the students' reactions and their progress on the initial phase were observed, there followed up with a curriculum neither too far above nor below the capacity and level of the average. Therefore, after the completion of two sessions of phase A, and with the foregoing factors in mind, phases B and C were formulated, with timing similar to that of phase A, as previously described; see Figs. 2 and 3.

The duration of any training course is most difficult to determine. Economy and available manpower play a part as important as training depth and needs. Sacrifices, sometimes extreme, must be made, as when, during the training period, the older engineers are obliged to isolate themselves from their families. Therefore, compromise is in order. In the



NYTEL COURSE E.E. 13

Term: 7 to 8 Weeks (1 class session per week)
Text: Kerchner & Corcoran:
"Alternating Current Circuits" - Wiley - 1950

Session	Topic	Text Assignment
1	Introductory concepts: Inst. Voltage, current & Power in R,L,C branches	PP. 1-11 12-36
2	Effective & average values - power - reactive power - p.f. simple phasor diagrams	PP. 42-62
3	Complex algebra-phasors - transmission constant of 2 port net	PP. 63-94
4	Single-phase circuits - resonance - dissipation factors	PP. 97-144
5	Coupled circuits - mutual imped. - Coeff. of coupling - magnetic coupling	PP. 211-238
6	Coupled circuit - singly & doubly tuned circuits	PP. 239-255
7	Supplementary problems	Notes

Fig. 6. EE 13 of NYTEL course

.78
1
59

case of the Bell System course, the weighing of the various factors (without benefit of an exact science) led to the plan of using three 3-week periods, or a total of 9 weeks. To absorb so much in so little time was not found easy by the students. Extensive homework was a necessity, and former college students recalled what it meant to "burn the midnight oil."

It may be well to relate the means taken to prepare certain of the prospective students for this curriculum. As mentioned earlier, preliminary examination of the students indicated that some of them required preliminary training to qualify for the course. Subjects planned for phase A and tentatively for phases B and C were reviewed and compared with the test results. There was sufficient indication of weakness in mathematics to warrant the development of two courses, Mathematics 11 and 12, the first a prerequisite for the second; see Figs. 4 and 5. Also, EE 13, an advanced course in a-c circuits was planned; see Fig. 6. These courses were designed for an instruction period of 3 hours, 1 day a week, for 7 to 8 weeks depending on the capabilities of each class. To reduce the time and the expense of traveling throughout the state, separate classes were organized as required, in the larger cities and towns, and instruction was given by a faculty member of the local college or university. With such competent instructors, the results were very favorable. In fact, experience has shown that many of the students who participated surpassed others in the Communications School who, based on the pre-examinations, were considered exempt from this preliminary training.

The selection of a faculty and a location for the new arts training did not prove to be very easy. Inquiry at several schools indicated an extreme shortage of instructors and, because of this, talent within the telephone industry was sought. It was immediately evident that active employees, competent in the skills and techniques of the new arts, were so engrossed in current problems that they also were unavailable. However, there is the never-ending flow of retiring engineers, particularly from the Bell Laboratories; experienced men with a wealth of knowledge, and the determination to remain active. This provided a source of supply from which teaching material could be readily drawn. Fortified with the required number of such talented men, the program was found to be acceptable by the School of Electrical Engineering at Cornell University. By accepting these men on its faculty and supplementing them with a few of its regular communications professors, the university has organized for the Bell System a well co-ordinated new arts course.

Living accommodations take on the collegiate atmosphere. The students taking the course live on the campus and, for sociability and group study, have been provided with space on the third and fourth floors of one of the smaller dormitories for their own use. Meals are served to the group as a whole in a private dining room, also located on the campus. Classes, and computation and laboratory periods, are held in Phillips Hall, the Electrical Engineering building. As previously mentioned, there are 24 students attending each session; actually, to obtain increased student participation and closer instructor-student relationship, a smaller number would have been preferred. However, the total number to be trained within the time limit required this larger group. To sample the progress of the students, a written examination was scheduled every Friday afternoon.

An important adjunct to the program was a special appreciation course for engineering management so that management, could become familiar with what their engineers were being taught. This course consisted of a 2-day combined lecture and laboratory session for the supervisors. It was designed not only to meet the requirement just mentioned but also to impart some general knowledge relating to the subjects taught. Such an Appreciation course should be mandatory in any training course plan.

The tasks of screening the engineers by pre-examination, providing advance training to those who needed it, and selecting the curriculum, the instructors, and the college were accomplished not necessarily in the order given nor independently. Rather, all were undertaken practically simultaneously, relating as they did to a closely integrated program. Many compromises were made, such as those in connection with the period of training and the size of classes. However, the company has the compensation of a well-coordinated educational program with sufficient emphasis on telephony so that the engineers, in living up to the tradition of this industry, may plan for better service at lower costs. The success of the program provides a striking example of cooperation between businessmen and educators for the common cause of technological improvement and advancement.

◆

Discussion

Ellis F. King (University of California, Los Angeles, Calif.): Mr. Waldner explains why industry must provide additional schooling for its engineers. Perhaps the term "education" is preferable as opposed to "training" for programs of this nature, but no fault is found with his conclusions.

Indeed, important additional evidence is found of fundamental changes in higher education patterns which reinforce his argument for increased industry-sponsored programs. Paradoxically, future engineering graduates will be both better educated and more in need of courses such as the ones described.

The trend in major engineering schools is toward more fundamental concepts at increasing levels of sophistication. The graduate of the future will be more broadly educated at the expense of knowledge of the techniques of a particular industry. And rightly so, for it is here that the colleges and universities can make their optimum contribution to the individual's eventual development to professional stature. For example, the advent of the transistor greatly reduced the value of a body of preceptual knowledge concerning vacuum-tube parameters and techniques, leaving undiminished (or even augmented) the need for a command of basic circuit analysis.

Carrying this thesis a bit further, imagine an engineer educated without regard to specialization in any of the traditional fields, but soundly competent in such fundamental modes of engineering analysis as:

1. Linear circuit theory (applicable to certain cases of heat transfer, fluid flow, and mechanical and acoustical devices, as well as electrical circuits).
2. Nonlinear analysis, including piecewise linear approximations and graphical methods.
3. Field theory, including experimental, graphical, and numerical solutions.
4. Statistical methods.

It is not difficult to visualize an industry program which could build upon such foundations the necessary techniques and application information to produce engineers for effective productivity in the telephone industry or in any industrial environment.

If one recognizes the existence of a long-range trend toward emphasis on fundamental principles and concepts in engineering education (and there is increasing evidence of the acceptance of this philosophy) then industry must indeed assume additional responsibilities in staff development.

Application of Finite Integral Transforms to Analog Simulations

WALTER J. KARPLUS
NONMEMBER AIEE

PREWITT A. STEPHENS, JR.
NONMEMBER AIEE

INTEGRAL transformations are defined by the expression

$$(p)=\int_a^b f(x)K(p,x)dx \qquad (1)$$

where $K(p, x)$ is termed the transformation kernel. Theoretically a wide variety of functions could serve as kernels, and an infinite number of transformations could be generated. However, only a limited number of kernels possess those properties which make their transforms useful for the solution of practical problems.

Most engineers are familiar with the Laplace transform whose kernel is ϵ^{-px} where $p=\sigma+jw$. In this case the limits on the integral are zero and infinity. This transform is, therefore, useful for the solution of problems in a semi-infinite region. Since the time domain in engineering problems has this characteristic, Laplace transforms are particularly useful for the solution of ordinary differential equations in which time is the independent variable.

In treating field problems characterized by partial differential equations, finite integral transforms comprise a valuable tool. As described by Tranter,[1] finite integral transformations are those transforms for which the limits a and b of equation 1 remain finite. These limits are chosen so as to correspond to the co-ordinates of the field boundaries along the independent space variable being transformed. As shown in the Appendix, the kernel $K(p, x)$ can be any function belonging to an orthogonal set. The kernel suitable for any specific problem is determined by the form of the analytical solution of the problem in the co-ordinate system being used. The analytical technique of separation of variables, applicable to 11 different co-ordinate systems,[2] leads to ordinary differential equations whose solutions are usually infinite series.

These infinite series comprise the kernel functions for the construction of the transforms. Thus, for example, in Cartesian co-ordinates sine and cosine transforms are employed; in cylindrical co-ordinates Hankel transforms are used; and the Legendre transforms are applicable to spherical co-ordinates.

Analytical methods for solving field problems are limited to fields having relatively simple geometries. To analyze fields with irregular boundaries, numerical or analog methods must be employed. In this paper the application of integral transforms to facilitate analog simulation methods is considered.

The Finite Sine Transform

When a function $f(x)$, $0\leq x\leq\pi$, can be expressed as the sum of sinusoidal terms in the well-known Fourier Series

$$f(x)=\frac{2}{\pi}\sum_{p=1}^{\infty}\bar{f}(p)\sin(px),\ 0\leq x\leq\pi \qquad (2)$$

then the coefficient $\bar{f}(p)$ of the p th term defines the finite sine transform, where

$$\bar{f}(p)=\int_0^\pi f(x)\sin(px)dx \qquad (3)$$

Equations 2 and 3 are the sine transform pair where equation 3 defines the transform and 2 its inverse.

To find the transform of $d^2\phi/dx^2$

$$\bar{f}(p)=\int_0^\pi d^2\phi/dx^2\sin(px)dx \qquad (4)$$

Integration by parts of equation 4 leads to

$$\bar{f}(p)=[\sin(px)(d\phi/dx)]_0^\pi - p\int_0^\pi d\phi/dx\cos(px)dx \qquad (5)$$

Paper 59-10, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted July 3, 1958; made available for printing November 10, 1958.

WALTER J. KARPLUS is with the University of California, Los Angeles, Calif., and PREWITT A. STEPHENS, JR. is with Hughes Aircraft Company, Culver City, Calif.

Vol.
Pt. 1
1959

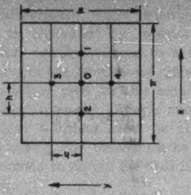


Fig. 1. Finite difference grid in two dimensions

The first term on the right-hand side of equation 5 is zero if $d\phi/dx$ is finite in the interval $0 \leq x \leq \pi$, and further integration by parts leads to

$$(p) = -p[\phi(x)\cos(px)]_0^\pi - p^2\int_0^\pi \phi(x)\sin(px)dx \qquad (6)$$

Now define $\bar{\phi}(p)$ as

$$\bar{\phi}(p) = \int_0^\pi \phi(x)\sin(px)dx \qquad (7)$$

Evaluating equation 6 and substituting equation 7 leads to

$$\bar{f}(p) = p[\phi_{x=0} - (-1)^p \phi_{x=\pi}] - p^2\bar{\phi}(p) \qquad (8)$$

Equation 8 shows that $\phi(x)$ must be known at $x=0$ and $x=\pi$. These are the specified boundary conditions of the problem. If the derivatives are specified at the boundaries, the cosine transform is used instead of the sine transform. If $\phi(x)$ is specified at $x=0$ and $d\phi/dx$ at $x=\pi$, then a special sine transformation is needed.

The Application of the Sine Transform

As a simple demonstration of the use of the sine transform, consider the two-dimensional heat transfer in a rectangular plate. The field equation to be solved is

$$\partial^2 V/\partial x^2 + \partial^2 V/\partial y^2 = (1/k)\partial V/\partial t \qquad (9)$$

The boundary conditions for simplicity are taken to be

$$V=0;\ x=0,\ \pi,\ y=0,\ \pi,\ t>0 \qquad (10)$$

For fields of other dimensions, a simple linear transformation can be employed. The initial condition is taken to be

$$V=V_0;\ t=0,\ 0 \leq x \leq \pi,\ 0 \leq y \leq \pi \qquad (11)$$

First, V is transformed to $\bar{V}$ by means of equations 7, 8, and 10 in order to eliminate $\partial^2 V/\partial y^2$ from the field equation. This leads to

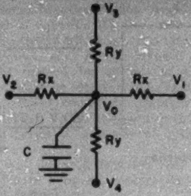


Fig. 2. Typical node of R-C analog for diffusion equation

$$\partial^2 \bar{V}/\partial x^2 - p^2\bar{V} = (1/k)\partial\bar{V}/\partial t \qquad (12)$$

Transforming the initial condition of equation 11 leads to

$$\bar{V}_0 = \bar{V}(x,\ p,\ 0) = \int_0^\pi V_0 \sin(py)dy, \quad t=0,\ 0 \leq y \leq \pi \qquad (13)$$

and from equation 13 the transformed initial condition is obtained:

$$\bar{V}_0 = \begin{cases} 0,\ p=\text{even} \\ 2V_0/p,\ p=\text{odd} \end{cases} \qquad (14)$$

Application of the sine transform to equation 9 has resulted in a new partial differential equation with new initial conditions. The boundary conditions on y could also be transformed, but since they are specified as being zero, they remain zero.

The sine transform is now applied to $\bar{V}'$ to eliminate $\partial^2\bar{V}/\partial x^2$ from equation 12. This leads to

$$d\bar{V}'/dt + (p'^2+p^2)k\bar{V}' = 0,\ p=\text{odd} \qquad (15)$$

where $\bar{V}$ is the doubly transformed variable and p' is transform of x. The initial condition for this equation is found as in equations 13 and 14 and is

$$\bar{V}_0' = 4V_0/pp',\ t=0,\ p,\ p'=\text{odd} \qquad (16)$$

The double application of the sine transformation to equation 9 has reduced this equation to the ordinary differential equation, equation 15. The solution of equation 15 is readily obtained as,

$$\bar{V}' = \frac{4V_0\epsilon^{-n^2kt}}{pp'}\ p,\ p'=1,\ 3,\ 5,\ \ldots. \qquad (17)$$

where,

$$n^2 = p'^2 + p^2 \qquad (18)$$

The final solution is obtained by applying the inverse sine transform equation 2 to equation 17 twice. This leads to,

$$V = \frac{16V_0}{\pi^2}\sum_{p=1}^{\infty}\sum_{p'=1}^{\infty}\frac{\epsilon^{-n^2kt}}{pp'}\sin py \sin p'x \qquad (19)$$

$p,\ p'=1,\ 3,\ 5,\ \ldots$

Analog Solution of the Untransformed Diffusion Equation

The electric analog solution of the diffusion equation 9 can be effected in several ways.[3] Here we will consider the conventional R-C (resistance-capacitance) network analog. The problem defined by equations 9, 10, and 11 may be represented by a 2-dimensional mesh where the spacing between nodes is designated as h.

For any particular node such as node zero in Fig. 1, equation 9 is expressed in finite difference form. This allows one to write the partial derivatives of equation 9 as differences in the temperature T between nodes 1-0, 3-0, etc. So, equation 9 becomes

$$\frac{(T_1-T_0)}{h^2} + \frac{(T_2-T_0)}{h^2} + \frac{(T_3-T_0)}{h^2} + \frac{(T_4-T_0)}{h^2} = \frac{1}{k}\frac{dT_0}{dt} \qquad (20)$$

If temperature T is made analogous to voltage V, h^2 corresponds to resistance, and $(1/k)$ corresponds to capacitance. Equation 20 then becomes

$$\frac{V_1-V_0}{R_x} + \frac{V_2-V_0}{R_x} + \frac{V_3-V_0}{R_y} + \frac{V_4-V_0}{R_y} = C\frac{dV_0}{dt} \qquad (21)$$

where $R_x = R_y = h^2$; $C = 1/k$. Equation 21 is recognized as the Kirchoff's current law equation for a junction of five elements as shown in Fig. 2.

Analog Solution of the Singly Transformed Diffusion Equation

The electrical analog of the transformed equation 12 is constructed in a similar manner to that described previously.

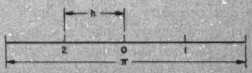


Fig. 3. Finite difference grid in one dimension

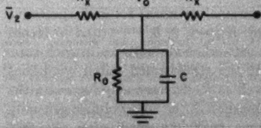


Fig. 4. Typical node of R-C analog for transformed diffusion equation

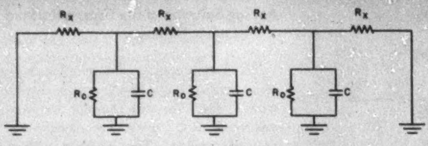


Fig. 5. Analog network for transformed diffusion equation

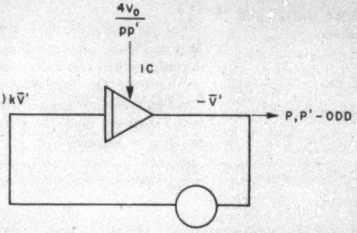


Fig. 6. Electronic analog for doubly transformed diffusion equation

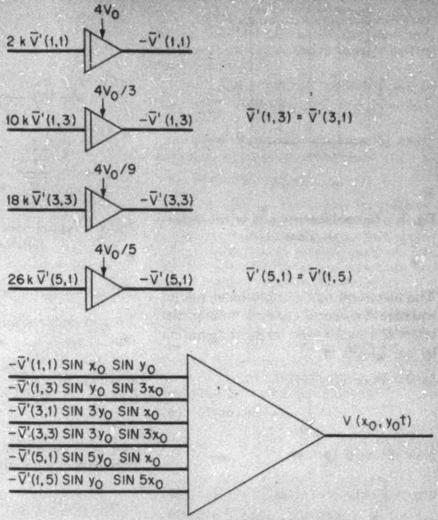

$4V_0$
$2k\bar{V}'(1,1)$
$-\bar{V}'(1,1)$
$4V_0/3$
$10k\bar{V}'(1,3)$
$-\bar{V}'(1,3)$
$\bar{V}'(1,3)=\bar{V}'(3,1)$
$4V_0/9$
$18k\bar{V}'(3,3)$
$-\bar{V}'(3,3)$
$4V_0/5$
$26k\bar{V}'(5,1)$
$-\bar{V}'(5,1)$
$\bar{V}'(5,1)=\bar{V}'(1,5)$
$-\bar{V}'(1,1)$ SIN x_0 SIN y_0
$-\bar{V}'(1,3)$ SIN y_0 SIN $3x_0$
$-\bar{V}'(3,1)$ SIN $3y_0$ SIN x_0
$-\bar{V}'(3,3)$ SIN $3y_0$ SIN $3x_0$
$-\bar{V}'(5,1)$ SIN $5y_0$ SIN x_0
$-\bar{V}'(1,5)$ SIN y_0 SIN $5x_0$
$V(x_0, y_0 t)$

Fig. 7. Electronic analog system for automatic inversion

Vol.
Pt. 1
195

However, this equation has only one space variable x. In finite difference form, equation 12 becomes

$$\frac{(\bar{V}_1-\bar{V}_0)}{h^2}+\frac{(\bar{V}_2-\bar{V}_0)}{h^2}+\frac{(-\bar{V}_0)}{1/p^2}=\frac{1}{k}\frac{d\bar{V}_0}{dt} \qquad (22)$$

where the subscripts refer to the line elements shown in Fig. 3.

The network section for a typical node is derived from equation 22 and is shown in Fig. 4, where

$$R_x=h^2,\ R_0=1/p^2 \qquad (23)$$

and

$$C=1/k \qquad (24)$$

If the network sections of Fig. 4 are placed at each node of Fig. 3; if the two end points are grounded to correspond to the zero boundary conditions; and if the initial conditions are applied as a charge at each node, then one obtains the network shown in Fig. 5 which is the electric analog of equation 12. This network can be used in lieu of the conventional two-dimensional R-C network.

According to equation 23, R_0 is a function of p. It is therefore necessary to construct a new network (or to modify the original network) for each successive value of $p=1, 3, 5, \ldots$ and to record the network node voltages in each case. Since the transformed initial conditions as well as R_0 decrease as p increases, the magnitudes of the node voltages generally become negligibly small after the first few values of p.

To obtain $V(x, y, t)$ from the analog solution $\bar{V}(x, p, t)$ one must use the following inversion equation:

$$V(x, y, t)=\frac{2}{\pi}\sum_{p=1}^{\infty}\bar{V}(x, p, t)\sin(py)\ p\text{–odd} \qquad (24A)$$

This inversion process may be done analytically or by using the analog technique described in the following section.

Analog Solution of the Doubly Transformed Diffusion Equation

Equations 15 and 16 imply the solution of an ordinary differential equation. The two successive applications of the sine transform have entirely eliminated the space variables. The two-dimensional diffusion equation therefore, can be solved by an electronic differential analyzer using only one integrator as shown in Fig. 6. The output of this circuit is proportional to $\bar{V}'$. The doubly transformed $\bar{V}'$ must then be inverted to find the temperature V, according to

$$V(x, y, t)=\frac{4}{\pi^2}\sum_{p=1}^{\infty}\sum_{p'=1}^{\infty}\bar{V}'(p, p', t)\times \sin(py)\sin(p'x) \qquad (25)$$

where p, p' are both odd integers.

The summation required by equation 25 may be performed on an electronic analog computer. If a separate integrator is used for each combination of p and p', and if all their outputs are applied to a summing circuit after appropriately scaling each output by $\sin(py_0)\sin(p'x_0)$, the output of the summer is $V(x_0, y_0, t)$. Such a circuit is shown in Fig. 7. Since both p and p' range from one to infinity, it is obviously impossible to treat all combinations of these parameters. However, since the initial conditions on the integrator are divided by pp', the voltage outputs of the integrators rapidly converge to zero as either p or p' becomes large.

This summing circuit only produces a solution at one point (x_0, y_0). A similar circuit would be needed for every other point desired. One could also reset the parameters and make additional runs to obtain additional points. This requirement of extra equipment or additional runs tends to offset some of the advantages of this technique.

Conclusions

The application of the sine transform to a simple field problem governed by the diffusion equation has been demonstrated in the foregoing. The method is readily extended to other partial differential equations and other co-ordinate systems. From the point of view of electric analog simulation, this method effects a considerable saving in equipment. A three-dimensional system requiring hundreds or thousands of resistors and capacitors can be reduced to a two-dimensional or one-dimensional system. A price must be paid for this economy in equipment, however. A large number of parametric solutions must be obtained on the analog, and these must be scaled and summed to obtain the final inverted solution.

Appendix

The General Finite Integral Transformation

Since a finite integral transformation is derived from an infinite series of orthogonal functions, it is possible to define a general expression for a finite integral transformation in terms of a general orthogonal function. After this general expression is derived, the well-known transforms such as the sine and cosine transform can be obtained readily.

Let us define an infinite series of functions

$$\phi_1(x),\ \phi_2(x),\ \ldots\ \phi_n(x),\ \ldots \tag{26}$$

which has the following properties

$$\int_a^b \rho(x)\phi_m(x)\phi_n(x)dx \begin{cases} =0, & m\neq n \\ \neq 0, & m=n \end{cases} \tag{27}$$

If the series equation 26 has the properties specified in equation 27, the series is said to be orthogonal with respect to the weighting function, $\rho(x)$ in the interval (a, b). A series defined by $\sin(nx)$ or $\cos(nx)$, where n takes on integer values and $\rho(x)$ is unity, satisfies condition equation 27 in the interval 0, π. So the well-known Fourier Series is a series of orthogonal functions.

One can define a function $f(x)$ in the interval a to b by an appropriate infinite series as follows

$$f(x)=\sum_p A_p\phi_p(x),\ a\leq x\leq b \tag{28}$$

In order to determine A_p, multiply equation 28 by $\rho(x)\phi_q(x)$ and integrate over the interval a, b.

$$\int_a^b \rho(x)f(x)\phi_q(x)dx = \sum_p A_p \int_a^b \rho(x)\phi_p(x)\phi_q(x)dx \tag{29}$$

From equation 27 it is seen that all terms are zero except when $p=q$. So, equation 29 becomes

$$A_p=\frac{\int_a^b \rho(x)f(x)\phi_p(x)dx}{\int_a^b \rho(x)\phi_p^2(x)dx} \tag{30}$$

Substitution of equation 30 into equation 28 leads to,

$$f(x)=\sum_p \frac{\phi_p(x)\int_a^b \rho(x')f(x')\phi_p(x')dx'}{\int_a^b \rho(x')\phi_p^2(x')dx'} \tag{31}$$

where $a\leq x\leq b$.

Now, if the numerator of equation 30 is defined as $\bar{f}(p)$, and inserted in equation 31, one obtains the following transform pair:

$$\bar{f}(p)=\int_a^b \rho(x)f(x)\phi_p(x)dx \tag{32}$$

$$f(x)=\sum_p \frac{\bar{f}(p)\phi_p(x)}{\int_a^b \rho(x')\phi_p^2(x')dx'},\ a\leq x\leq b \tag{33}$$

Equations 32 and 33 are the general expressions defining a finite integral transformation pair.

References

1. INTEGRAL TRANSFORMS IN MATHEMATICAL PHYSICS (book), C. J. Tranter. John Wiley & Sons, New York, N. Y., 1951.

2. METHODS OF THEORETICAL PHYSICS (book), P. M. Morse, H. Feshbach. McGraw-Hill Book Company, Inc., New York, N. Y., 1953.

3. ANALOG SIMULATION: SOLUTION OF FIELD PROBLEMS (book), W. J. Karplus. McGraw-Hill Book Company, Inc., 1958.

Computational and Experimental Techniques in Nuclear Reactor Design

W. F. WITZIG
MEMBER AIEE

M. R. STUART
NONMEMBER AIEE

L. O. HERWIG
NONMEMBER AIEE

THE PURPOSE of this paper is to discuss a nuclear design technique for a nuclear power reactor. It is felt that this area of reactor design is not as well-known in the engineering profession as, for example, power plant systems or control systems.

The design of a nuclear power plant begins with a list of technical objectives which are arbitrarily divided into two groups, engineering design and nuclear design, as shown in Fig. 1. Technical objectives include such items as size, power capacity, core life, methods of reactivity control, materials, and other items such as operating temperatures and temperature coefficients of reactivity that are more closely related to the plant system.

For the purposes of this discussion, only the right-hand branch of Fig. 1, the nuclear design, will be considered. It has been divided into two activities: 1. experimental, and 2. analytical. The experimental activity involves a critical facility and its associated program. The experimental facility is described in a paper by E. S. Lembersky.[1]

To help establish what is meant by nuclear design, it is necessary to resort to an illustration of the design sequence; see Fig. 2. This chart not only shows the sequence of events, but helps establish the relationship between the analytical and the experimental activities. Knowing the nuclear core objectives, the next step is to develop an elementary design model. Such a model is generally very simple; the core inventory is assumed to be homogeneously distributed throughout the volume. Employing this model, it is customary to study the reactivity or effective neutron multiplication of various core inventories.

The inventory variables include the quantity of fissionable material, structure, moderator, and perhaps more important, the quantity of poison or control material. Evaluation of the results leads to a preliminary design which includes such items as the distribution of the nuclear materials, the geometry, the number and the distribution of the reactor control rods, the size of the reactor core, and the metal-to-water ratio. The preliminary nuclear design is closely related to the engineering design, as indicated in Fig. 1. However, this discussion concerns only the nuclear aspects of this design.

At this point both the experimental and the analytical activities can proceed in parallel. Close liaison is required between these two efforts.

Paper 58-1315, recommended by the AIEE Nucleonics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26-31, 1958, and re-presented at the Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted August 25, 1958; made available for printing December 3, 1958.

W. F. WITZIG, M. R. STUART, and L. O. HERWIG are with the Westinghouse Electric Corporation, Pittsburgh, Pa.

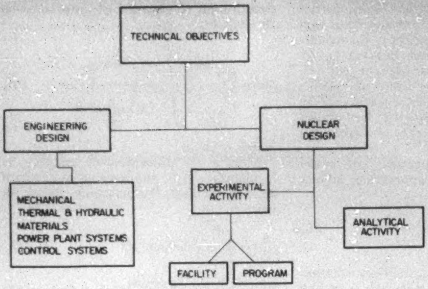

TECHNICAL OBJECTIVES
ENGINEERING DESIGN
NUCLEAR DESIGN
MECHANICAL
THERMAL & HYDRAULIC
MATERIALS
POWER PLANT SYSTEMS
CONTROL SYSTEMS
EXPERIMENTAL ACTIVITY
ANALYTICAL ACTIVITY
FACILITY
PROGRAM

Fig. 1. Division of technical responsibility

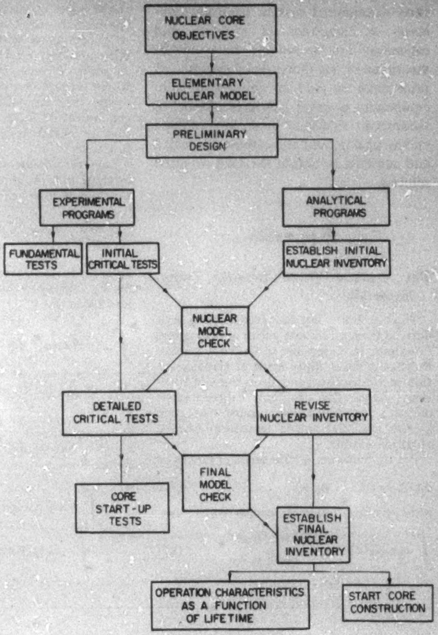

NUCLEAR CORE OBJECTIVES
ELEMENTARY NUCLEAR MODEL
PRELIMINARY DESIGN
EXPERIMENTAL PROGRAMS
ANALYTICAL PROGRAMS
FUNDAMENTAL TESTS
INITIAL CRITICAL TESTS
ESTABLISH INITIAL NUCLEAR INVENTORY
NUCLEAR MODEL CHECK
DETAILED CRITICAL TESTS
REVISE NUCLEAR INVENTORY
FINAL MODEL CHECK
CORE START-UP TESTS
ESTABLISH FINAL NUCLEAR INVENTORY
OPERATION CHARACTERISTICS AS A FUNCTION OF LIFETIME
START CORE CONSTRUCTION

Fig. 2 (right). Nuclear design sequence

With a somewhat limited ability to predict criticality of the reactor the primary emphasis of the experimental program is to provide measurements which can be compared to calculations based on the nuclear design model. Criticality is the characteristic of a reactor core which describes the core's ability to sustain a nuclear chain reaction. In the general case the experimental program is directed at studies of the critical assembly, which are described in the section dealing with "Experimental Methods in Reactor Design." In addition to the uncertainties of the nuclear model for predicting criticality of this complicated assembly of nuclear materials, there may be uncertainties in the values of the constants which are used in the model. Examples of some of these constants are the neutron cross section of the products generated as a result of the fission process, the resonance absorption in U-238 which leads to the production of plutonium, and the product of the xenon fission yield and neutron cross section. A portion of the experimental program is often devoted to measuring the quantities associated with the calculation of reactor constants.

In parallel with the experimental effort the analytical work proceeds, using a model more nearly representative of the reactor core than the early elementary nuclear model. Such a model begins to take into account (with two-dimensional analysis) the core heterogeneities, the geometry of control rods, and the distribution of nuclear fuel.

Having at this stage both experimental and analytical work of sufficient detail, a check of the nuclear model is performed. Then it is possible to refine the nuclear model so it is a more exact representation of the power reactor. These refinements then lead to both detailed experimental programs and analytical programs, which are again woven into a final model check. At this point we then can move from the final model check into a final determination of the nuclear inventory.

It is desirable to continue the experimental work on what might be called core start-up tests, which involve such items as biological shielding studies, and neutron source strengths for start-up. If there has been some operating experience with this type of reactor, it is possible at the final nuclear inventory stage to initiate core construction. However, the operating characteristics as a function of lifetime of the core may well be predicted prior to the initiation of core construction. Examples of these operating characteristics are the criticality of the core as a function of lifetime and the power distributions throughout this lifetime.

In the second section of this paper the nuclear model is examined in greater detail. In the third section the computation techniques used in the analytical program to perform nuclear model checks and to give the engineering designer his required information are examined. In the fourth section, the experimental program is reviewed, showing typical experimental techniques involved in producing experimental data. In the fifth section of the paper, some thoughts about the future role of the computational and experimental techniques in reactor design are discussed.

Description of a Nuclear Design Model

The discussion thus far has been sufficient to point out that a nuclear reactor has a variety of characteristics, not all of which contribute to the peace of mind of the reactor designer. Some of the more objectional ones we have to live with, others we are able to circumvent, while it is possible to exploit still others to meet design objectives. The purpose of this section is to pull these characteristics together in a conceptual picture, which we choose to call the nuclear design model. Every designer, whether he be a mechanical, thermal, or hydraulic engineer or a physicist, has in his mind a picture of the reactor. More

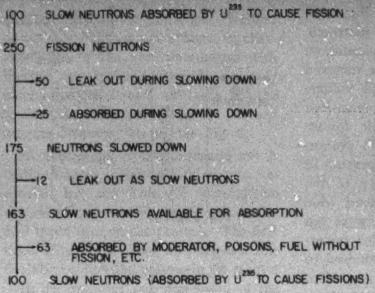

100 SLOW NEUTRONS ABSORBED BY U^{235} TO CAUSE FISSION
250 FISSION NEUTRONS
50 LEAK OUT DURING SLOWING DOWN
25 ABSORBED DURING SLOWING DOWN
175 NEUTRONS SLOWED DOWN
12 LEAK OUT AS SLOW NEUTRONS
163 SLOW NEUTRONS AVAILABLE FOR ABSORPTION
63 ABSORBED BY MODERATOR, POISONS, FUEL WITHOUT FISSION, ETC.
100 SLOW NEUTRONS (ABSORBED BY U^{235} TO CAUSE FISSIONS)

Fig. 3 (left). Typical events in the life of a neutron

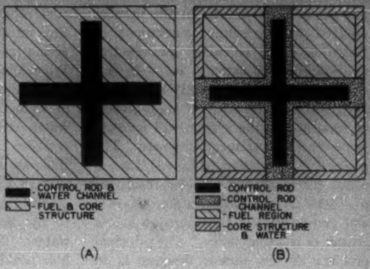


Fig. 4 (right). Core heterogeneities

A — Preliminary representation
B — Detailed representation

specifically he thinks of the particular aspects of the reactor in terms of the characteristics and qualities for which he is responsible. The mechanical engineer, for example, may view the reactor as a series of masses, spring-connected, each with a certain size, strength, and natural frequency. With this picture, or model, he then proceeds to evaluate its characteristics through a mathematical representation, which allows quantitative evaluations to be made and specifications to be determined. The quantitative results are then compared with the objectives of the design. Most frequently, the peculiar requirements of a given reactor design call for devices which are developmental in nature. Thus the engineer must supplement his calculations with tests involving mock-ups or models of the proposed design. The discussion here, though elementary, is to remind the reader that even today, with a discipline as old as that of mechanical engineering, often a semi-empirical approach is necessary in reactor design. This is due to in part to the stringent requirements of reliability but also to the lack of accumulated engineering experience in this rather new field.

NEUTRONS, ATOMS, AND FISSION

As one moves to the reactor physics aspect of reactor design, a situation similar to that of the mechanical designer is found. From the physics or nuclear design standpoint the reactor is viewed from a different conceptual picture, but the problems of design are of comparable difficulty. One must design a core which is producing heat by a nuclear reaction involving fissionable material, supported by structural material in a moderating environment, and permeated by a cloud of neutrons. Neutrons are infinitesimal particles of matter which are able to penetrate large thicknesses of ordinary material. The presence of neutrons in the nuclei of almost all materials contribute to their stability. With specific nuclei, however, an excess of neutrons can exert very disruptive forces. These forces cause the nuclei to fission, i.e., split into parts having large kinetic energy, which is shared with the environment. The fundamental problem of the reactor physics designer is to bring together many neutrons and fuel nuclei in such a manner as to cause a controlled release of this energy.

NEUTRON CYCLE

Consider a single fission neutron from its birth to its death (absorption) or escape from a reactor core. A fission neutron is born at a relatively large average energy (about 2 million electron volts) and is called a fast neutron. It has an average speed of about 10,000 miles per second. As the neutron moves swiftly through the core, it collides with atoms of the material in the reactor. These collisions may scatter or absorb the neutron. In a water-cooled reactor one of the principal neutron scattering processes is that with the hydrogen in the water. In this process the neutron loses some of its energy to the hydrogen atom at each collision. This sequence of energy loss is called the slowing down or moderation of the neutrons. Eventually, a neutron loses most of its speed and thereby increases its chances for interaction with fuel nuclei, since it spends a longer time in the vicinity of the fuel nuclei when it passes by at a lower speed. The fact that a neutron is absorbed by fuel and the fission process initiated, is the basis for the chain reaction which makes nuclear power feasible.

The fuel usually splits into two fragments, each with considerable kinetic energy which manifests itself as heat. In addition to the fission fragments, approximately 2.5 neutrons are born (represented by the symbol ν) in the fission of U-235. Thus, in analogy with chemical processes, the neutron acts as a catalyst for the generation of heat, and, in addition, reproduces itself by a factor of 2.5.

Two categories of neutrons are identified with the fission process: 1. neutrons released at the instant of fission of a fuel atom, called prompt neutrons, and 2. neutrons emitted by certain radioactive fission fragments as long as several minutes after the fission fragments are created, called delayed neutrons. Of the approximately 2.5 neutrons born in each fission process, the delayed neutrons make up less than 1%. However, it is these neutrons which exert a damping action upon the kinetic response of a reactor and ease the control engineer's problems.

If one were to follow these fission neutrons as they slowed down, it would be found that some escape the reactor core entirely and some interact with materials other than fuel. These losses of neutrons do not contribute to the production of power. However, if one of these 2.5 fission neutrons eventually slows down and is absorbed in fuel to cause fission, the process is self-sustaining and is called stationary, since the total number of neutrons remains constant in time. Such a condition of the reactor is called the "critical state." Fig. 3 illustrates in a simplified manner the events occurring following the absorption of 100 slow neutrons in the fission process. This results in the release of approximately 250 fast neutrons, some of which in the process of slowing down are absorbed or escape the core. The remaining neutrons are then available to repeat the process.

The time from neutron birth through absorption to the re-emission of a neutron in a subsequent fission process is called the neutron generation time. This cycle occurs on the average of 10,000 to 100,000 times per second. The ratio of the number of neutrons released in the fission process at a given time to the number released in succeeding generations of fission neutrons is called the effective multiplication constant, k_{eff}. If k_{eff} equals 1, the system exactly reproduces itself and is stationary. If k_{eff} is greater or less than 1, the system will increase or decrease, respectively, in power. The percentage deviation, $(k_{eff}-1)/k_{eff}$ is called the reactivity ρ. The reactivity ρ is zero for the critical state $(k_{eff}=1)$ and greater or less than zero as k_{eff} is greater or less than unity. The principal problem of the nuclear designer is to control both the competition between the various core materials for the absorption of the neutrons and the neutron escape from the reactor.

NEUTRON FLUX

To determine quantitatively the nuclear characteristics and to obtain the design objectives of the reactor, a mathematical description is necessary. The neutrons in the core are treated in the same manner as a diffusing gas, with the additional feature that absorption and creation of the neutrons are taken into account. Consider an elementary volume of the core in a stationary system; the leakage plus the absorption in the elementary volume must equal the source of neutrons. Expressing this balance mathematically gives rise to a set of coupled differential equations. A solution to this set of equations, with the proper boundary conditions, represents the neutron flux as a function of position in and around the reactor core. This neutron flux, commonly represented by the symbol ϕ, is a quantitative measure of the neutrons present at a particular point in the reactor core at a given instant of time. With this knowledge it is possible to calculate the expected rate of energy release, biological shielding requirements, and control requirements, all of which are reactor design objectives.

HETEROGENEITIES

As mentioned previously, two features of the reactor which contribute to its operation are: 1. slowing down of fast or energetic neutrons and 2. the fission process itself with the accompanying release of extra neutrons. Since most of the slowing down takes place in the moderator, distribution of water in the reactor is of importance in calculating power distributions. For instance, there are usually large regions of the core in which control elements are physically moved. This is illustrated in Fig. 4(A). The rather simple geometry illustrates a cross-shaped control rod surrounded with core structure and fuel. In Fig. 4(B) the number of regions is increased, giving more detail to the representation. When the control elements are removed from the core a relatively large volume of water (the moderator in pressurized water reactors) occupies their former position. Consequently, unless precautions are taken, the power generated in fuel in the immediate vicinity of the control rod channels is increased (due to the slowing down properties of the water). This behavior is called water-hole peaking and must be minimized to reduce nonuniform power generation. The power generated in the fuel plates next to the control rod channels is frequently the limiting factor in meeting thermal design requirements.

DESCRIPTION OF THE NUCLEAR DESIGN MODEL

Now that a few details of the core design have been described, a model that may be used to investigate the characteristics of a reactor is discussed. Two questions shall be asked: 1. What is required from the model? 2. What must the designer specify to conduct the calculations for a given model? In answer to the first question, the following are needed: 1. a materials inventory which gives a specified lifetime, 2. the number, size, shape and effect of control rods, 3. spatial power distribution throughout the core lifetime, and 4. the effect of changes in temperature upon the neutron multiplication process. To obtain answers to the second question, a designer must specify: 1. the geometry and the mode of depletion of fuel and poisons in the core; 2. the number of groups of neutrons with specified energies; and 3. values of the nuclear parameters which are required by the mathematical description of the core. The last is referred to as the reactor constants.

All characteristics of the reactor core depend to some extent upon the distribution of fuel within the core volume. Consequently, during the depletion process, the characteristics of the core change. The principal problem of a nuclear designer is to calculate the changing distribution of the fuel and burnable poisons, and in turn, the changing characteristics of the core. However, for estimates of the core performance throughout lifetime, it is convenient in the preliminary studies to neglect the spatial variations of the fuel and carry out the calculations as if the fuel were uniformly distributed. This model of the core is the first of two general classifications of nuclear design models, namely, the spatially uniform depletion model. This model is used to obtain the initial nuclear inventory, as given in Fig. 2 in the nuclear design sequence. An illustration of the spatially uniform depletion model is shown in Fig. 5.

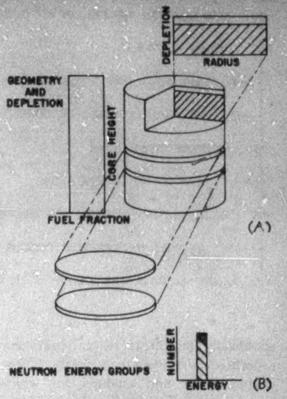


Fig. 5. Nuclear design of spatially uniform depletion model

Constants:*	
Absorption	Maxwellian
Fission	Maxwellian
Leakage	Constant

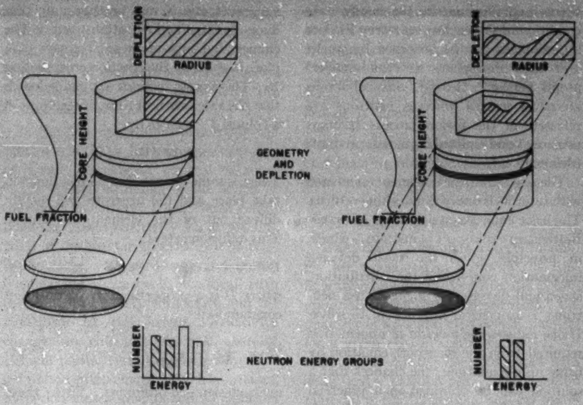
* For purposes of simplicity some liberty has been taken in assigning the type of constant to each model.

For the initial calculations to determine the preliminary estimate of the nuclear inventory, it is usually sufficient to assume a rather simple geometry for the core, that of a right circular cylinder, and in addition to assume that all materials (fuel, poisons, and core structure) are distributed homogeneously throughout this region. This is illustrated in Fig. 5(A). In addition, it is usually assumed that the calculations can proceed using the average properties of the group of neutrons whose energies are distributed like those of gas molecules in thermodynamic equilibrium at the particular temperature of the core. Therefore, this distribution of neutron energies is essentially a one-group representation. This is illustrated in Fig. 5(B). Consequently, the constants listed in Fig. 5 are calculated assuming the neutron energy distribution is the same as that of a gas at this particular temperature and, therefore, the Maxwellian averages are used. It should be pointed out that this model is only sufficient to determine an estimate of the materials inventory which will result in

Vol. 7
Pt. 1
1959

a given lifetime for the core. Neither power distributions nor changes in characteristics of the core parameters with lifetime can be calculated. In order to calculate these one proceeds in the nuclear design sequence to the revised nuclear inventory state; see Fig. 2.

At this point in the sequence use is made of some models which are characterized by the title of nonuniform depletion models. There are principally two types used at the present. The first type, Fig. 6(A), takes into account the depletion of fuel and poisons in the reactor with lifetime over one dimension of the core.[2] Usually this is in the axial direction. In this particular model the core constituents are still assumed to be distributed at the beginning of life in a homogeneous manner. Then the depletion calculations are initiated, and the core is divided for the purposes of calculation into a series of axial segments, somewhat similar to a stack of poker chips. The history of each of these chips is followed throughout the core lifetime, with the result that, as the depletion proceeds, these chips will contain differing amounts of fuel and poisons. In these studies, the energy distribution of the neutrons is extended to either two or four groups. With this extension, both in the spatial distribution of fuel and number of neutron energy groups, it is now possible to obtain estimates of the variation of certain properties of the core throughout lifetime. In addition to a more accurate estimate of the core lifetime for a given initial nuclear inventory, one can obtain axial power distributions and preliminary estimates of the control requirements necessary to meet the over-all design objectives. Also, one obtains reasonable estimates of the nuclear characteristics of the core with this one-dimensional depletion model, and information is obtained which is passed to the thermal and hydraulic engineers. These studies then result in a sufficient body of information to continue the nuclear design revisions. The principal advantage of this model is that it gives information sufficient for an estimate of the relative merits of several proposed designs, but is still simple enough to use in survey-type calculations.



(A) (B)

Fig. 6. Nuclear design of nonuniform depletion model

A—One-dimensional synthesis
B—Three-dimensional synthesis

Constants	Slow Neutron Group	Fast Neutron Group
Absorption	Maxwellian	Neutron Slowing Theory (reference 5)
Fission	Maxwellian	Fission permitted
Leakage		Varies with lifetime

When the final nuclear inventory is established the next stage of the design is to obtain the detailed operating characteristics of the core, both nuclear and engineering, as a function of its lifetime. This is done with a nonuniform depletion model, illustrated in Fig. 6(B). Use is again made of the concept of stacking horizontal segments of the core to obtain a nonuniform depletion of fuel throughout lifetime. However, the additional feature of the nonuniform depletion of the fuel and poisons throughout each axial segment is taken into account. The geometrical representation of the reactor in the horizontal plane is done in as much detail as possible, within the scope of the computing facilities. With these axial segments again superimposed, one obtains a three-dimensional representation of both power and fuel distributions throughout lifetime. This model, at the present time, is too complicated to use as a survey tool and is used only with the final nuclear inventory. To use this concept effectively, it is usually assumed that the neutron energy spectrum is composed of two energy groups. This assumption accounts for fission both in the fast and slow groups, with the resultant better estimate of the power distribution, and yet is simple enough to be used conveniently. In addition, one represents explicitly the control elements in the core and obtains a more accurate estimate of their effectiveness.

The preceding discussion outlined the various nuclear design models utilized as the design proceeds from the preliminary design, using a spatially uniform depletion model, to the final specifications of the core characteristics as a function of lifetime, using the three-dimensional superposition model. At each stage of the design sequence, the models are used to calculate the results of experiments performed with the critical assembly. These so-called "model checks" result in modifications of the nuclear inventory. The computing techniques for evaluating these models are discussed in the following section.

Computational Techniques

How physical concepts are fashioned into mathematical equations, and how solutions to these equations are obtained will now be indicated. The results of the analytic effort are phased with the experimental program to evaluate and develop the nuclear design model as indicated in Fig. 2 and further discussed in the section on "Experimental Techniques."

The main features of the mathematical treatment of reactors are taken directly from classical statistical mechanics, mainly the kinetic theory of gases. These include number density, mass flow, and the concept of scattering cross section. There are several distinct differences, however, between classical atoms in thermal equilibrium and the steady-state condition in the nuclear reactor. First of all, neutrons can be absorbed by nuclei and thus disappear, whereas classical atoms persist. Second, classical atoms are never created, whereas neutrons are released in the fission process (prompt neutrons) and emitted by fission products (delayed neutrons).

The mathematical treatment connected with reactors is one of diminishing returns. The basic equation, a linearized Maxwell-Boltzmann transport equation, which, in principle, gives the most detailed information concerning the distribution, speed and direction of flight of the neutrons, is in general impractical to solve. If one reduces the amount of information required from the transport equation by using approximations, the chances that a solution can be obtained in a practical manner are improved. It is, nevertheless, still necessary to utilize digital computers to solve these approximate equations in a reasonable length of time.

To illustrate some of the problems faced by the reactor designer, one of the main analytical problems is examined. If there are enough neutrons present in a reactor for statistical averages to be valid, then it is possible to use a point function, called the neutron flux, to calculate power distributions, flux peaking factors, and the variation of reactor control with changes in temperature and uranium depletion. For instance, the slow neutron flux would be that group of neutrons having speeds up to a certain upper limit, at a point within or about the reactor. Likewise, the second group of neutrons (fast groups) have speeds above the upper limit for the slow group.

The number of neutrons dn within and passing through a volume dV in a time dt is given by

$$dn = \phi/v\,dV\,dt \qquad (1)$$

where ϕ is the neutron flux and v is the average speed of the neutrons. If the reactor is operating at a steady rate, ϕ is only a function of position within the reactor. Strictly speaking, the flux ϕ is to be interpreted, after proper normalization, as the probability density of a neutron being present in a particular volume, whereas the actual number of neutrons present in a small volume fluctuates from time to time. If enough neutrons are present, however, these statistical fluctuations are very small.

As these neutrons pass through matter, there is a certain probability that various reactions take place between the neutrons and the atoms of matter present. The neutrons may be absorbed to produce very energetic X rays or they may produce fission in fuel atoms, etc. The chance that any of these processes take place is given by the reaction cross section Σ. These cross sections are defined such that the rate, per unit time, of reactions in a volume dV is given by

$$\text{Reaction rate} = \Sigma\phi\,dV \qquad (2)$$

There is some chance that no reaction will take place and the neutrons will merely diffuse out of the element of volume. This diffusion rate is:

$$\text{Diffusion rate} = -D\nabla^2\phi\,dV \qquad (3)$$

where D is a proportionality or diffusion constant and

$$\nabla^2\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} \qquad (4)$$

in cartesian co-ordinates. With these results, it is possible to write a representative equation in three dimensions, which exhibits the neutron economy under steady-state conditions.

$$(\underbrace{-D\nabla^2\phi} + \underbrace{\Sigma\phi} - \underbrace{S})\,dV = 0 \qquad (5)$$

$$\underbrace{\text{Leakage} + \text{Absorption}}_{\text{Loss}} = \text{Source}$$

In the two neutron group treatment there are two flux functions, ϕ_1 and ϕ_2, one for each neutron group. The equations which these flux functions satisfy are related, since the source for each group is proportional to the flux of the other group. The diffusion equations are

$$D_1\nabla^2\phi_1 - \Sigma_1\phi_1 + v\Sigma_2\phi_2 = 0 \qquad (6)$$

(thermal fission only)

$$D_2\nabla^2\phi_2 - \Sigma_2\phi_2 + \Sigma_1{}^R\phi_1 = 0 \qquad (7)$$

Equations 6 and 7 are approximated by a set of finite difference equations, with the result that a set of algebraic equations must be solved instead of a set of differential equations. This is done by numerical methods using various iterative schemes.[3] Since, in principle, an iterative scheme only converges after an infinite number of iterations, some error is introduced in the solutions by using only a finite number. The degree of convergence is controlled by the reactor designer on the basis of required precision, computing time, and cost.

There are at least three main areas of uncertainties in the analytic studies associated with reactor design: 1. the use of a group diffusion model instead of the more rigorous transport theory, 2. the fact that the reactor cross sections and diffusion coefficients may have large uncertainties, and 3. the presence of approximations in the numerical solution of the diffusion equations.

Examine a typical nuclear design problem in which it is desired to know the flux distribution along the radius of the core and compare the solution obtained by hand and by digital computer. A typical reactor design problem would involve something on the order of 90 to 100 points within the reactor and 40 to 50 points in the reflector at which it is desirable to find the flux. The solution to this problem requires something of the order of 10^8 mathematical manipulations to solve the difference equations. A girl at a desk-calculating machine can complete about six operations per minute. At this speed it would require something like 20 years to complete this problem, working 24 hours a day. By comparison, a modern high-speed digital computing machine would require only 10 minutes to solve the same problem, with a much higher probability of obtaining the correct answer.

An analogous problem exists in the study of load flow in an electric power network. Such a problem might involve a network having 100 nodes at which we wish to know the flow of real and reactive power. The solution of the simultaneous equations involved requires about the same time on a digital computer as the above nuclear design problem.

Experimental Methods in Reactor Design

Critical assemblies of reactor materials have played an important role in the nuclear design of pressurized water reactors. These assemblies serve to validate nuclear models and provide core startup data, measurements of reactor constants, and operating experience. Such assemblies may be arbitrarily classified as engineering mock-ups, clean criticals, and prototypes. An engineering mock-up is one whose nuclear characteristics correspond closely to the characteristics of the power reactor, except for operating conditions of temperature, pressure, and power. Construction is economical in time and costs, and core arrangement is flexible for study of parameter and design changes.[1] A clean critical assembly incorporates the same advantages in time, cost, and flexibility, and is specifically designed to measure particular reactor parameters with configurations closely applicable to the analytical model. These two types are designated as flexible critical assemblies. A prototype power reactor is an exact reproduction of the design and is quite inflexible with regard to changes in design. Experimental methods can be applied to each of these assemblies to obtain design information. An experimental method good for one may not be applicable for the other types. The discussion of experimental methods will deal largely with the use of a flexible critical assembly for the nuclear design of pressurized water plants.

The critical is constructed from small unitized core materials, and is made to look as much like the power reactor core as is practical. The nuclear design model is checked by this critical near room temperature, at low power, largely at the start of core life (no uranium depletion), and for a wide range of the core parameters.

Quantities which can be measured include critical mass of fuel materials, power distribution, temperature coefficient of reactivity, reactivities with control elements in various configurations, and effects of core and reflector materials. In addition, the critical assembly provides engineering information which includes studies of startup instrument response relative to startup neutron source strength, core startup characteristics, and biological shielding characteristics.

In addition to the flexible critical type of experiments, there are varieties of critical experiments directed at the understanding of the basic nuclear parameters incorporated in the nuclear design model. Some of the nuclear parameters of interest to the nuclear design model include measurements of the stable fission product cross section, xenon transient characteristics, energy distributions of neutrons in the core and the shielding around the core, activation of the shielding or coolant materials, and macroscopic cross section changes of fuel and poison as a function of depletion.

TWO PHILOSOPHIES IN THE OPERATION OF CRITICAL ASSEMBLIES

Reactivity changes in a critical assembly may arise from a large number of sources, e.g., movement of control elements and changes in fuel and poison loading or distribution, in temperature, in core metal-to-water ratio, in core or reflector structural materials, in core geometry, in power, etc. Two operational philosophies can be pointed out in the use of critical assemblies to determine reactivity changes during design studies. First, reactivity differences are measured in terms of position changes of calibrated control elements required to maintain the reactor critical, or in terms of the rate of change of the reactor power as a function of time (period measurements) following a calibrated control element position change. Second, the reactivity changes are measured in terms of a quantitative change of nuclear materials in the core to maintain a critical reactor with no change in the position of the control element.

In the first operational philosophy the control reactivity in many critical experiments consists of control elements in the referenced design or the height of the water moderator. Under these conditions the excess reactivity k_{ex} at the disposal of the reactor operator can be relatively large. Excess reactivity is defined as that above $k=1$ (or critical) with the control rods fully withdrawn and the moderator at full height. Reactivity changes are calculated from the control element position and the reactivity worth per unit position change, which is obtained by period measurements and an equation relating period to reactivity (the in-hour equation).

In the second operational philosophy, reactivity control in critical experiments often consists of auxiliary control rods whose effects are not part of the design study. The reactor is maintained critical by the auxiliary control rods, which are largely removed from the fuel region and whose reactivity effects are small and accurately known. An experimental change in core reactivity is measured in terms of the reactivity worth of a quantity of the nuclear material. This reactivity worth is measured by several methods, including period measurements or subcritical counting rate measurements. If only the auxiliary rods can be moved during operations at criticality, the excess reactivity at the disposal of the operator is only slightly greater than zero. The number of rod withdrawal mechanisms required for operations is small even for large cores with many design rods.

In the experimental methods to be discussed, the two operational philosophies have been combined to obtained design data. All the design rods can be moved during operations at criticality. However, part or all of the design rods can be designated as auxiliary rods, if desirable, and the amount of nuclear material adjusted to measure changes in reactivity. Occasionally, when all design rods are required to be all or partially inserted in the core for an experiment, auxiliary rods are added.

The following classification of experimental methods is chosen arbitrarily:

1. **Critical mass is that quantity of fuel and poison material, which, in conjunction with the reactor structure and moderator, is capable of maintaining a constant power level without an external source of neutrons.**
2. **Reactivity changes or changes in the neutron multiplication constant.**
3. **Neutron flux distribution shows the density of neutrons as a function of position throughout the volume of the core, and is called a power distribution if the fuel loading is uniform.**

Data concerning many reactor parameter may be obtained from a combination of these methods.

CRITICAL MASS METHODS

Many characteristics of a critical assembly are calculated in terms of the configuration and quantity of fuel and poison materials in the core. In particular, the quantity of nuclear fuel or posion material for a given critical configuration can be measured accurately and compared with analytical calculations. The poison material is that neutron absorbing material which is introduced into the reactor for reactivity control rather than for cooling or structural reasons.

One of the first checks of a nuclear design model is a comparison of the calculated and experimental critical masses for a rod-free critical assembly, a clean critical assembly. It is advantageous to work with a fuel element structure as nearly like the power reactor core as is possible, retaining flexibility for changes in nuclear inventory. The experiments may range from relatively simple and homogeneous structures to complex heterogeneous types. The models used to explain the characteristics must also incorporate these heterogeneities; see the section on "Description of a Nuclear Design Model."

The ability of the nuclear design model to include the effects of control rods can be checked by observing the critical mass as a function of control rod position in the core. Calculation of rod effects, particularly for cores containing large numbers of control rods, are simplified by obtaining critical masses of fuel and poison for combinations of design rods either fully in or fully withdrawn from the core. Measurement of the critical mass is an example of the second operational philosophy. The experimental test is a form of a null measurement to obtain $k_{ex}=0$.

REACTIVITY CHANGES

Critical mass measurements require equilibrium power conditions, whereas the reactivity measurements depend on transient behavior of reactor power (reactor period). An equation known as the in-hour equation relates the period and reactivity. The reactivity calculated from this equation depends upon the yield, the lifetime, and the effectiveness of delayed neutrons relative to the prompt neutrons. There are variations in delayed neutron parameters among nuclear fuels, and the effectiveness also varies with reactor configuration. Corrections for these variations can be calculated and applied to the experimental data.

Period measurements are used to determine the differential reactivity worths of control elements, fuel poison, and moderator height. Also, period measurements are basic to measuring changes in reactivity worth induced by placing a sample of material in a reactor core and by changing reactor temperature.

The differential reactivity worths of control rods and moderator height are obtained as follows. First, the reactor power level is stabilized or the period is determined for a given reactor configuration. Next, the control rod or moderator height is changed an accurately determined amount while holding other reactor parameters constant. A new period is then measured. Knowing the periods, the associated reactivity change can be calculated. The differential worth may be found by dividing the change in reactivity by the change in height, thus calibrating the control rod or moderator.

The reactivity worth of fuel or poison material is obtained by altering the material in the core by a known amount and observing the change in critical position of a calibrated control rod or in a reactor period. The temperature coefficient is measured by noting the position of a calibrated rod or the reactor period as a function of reactor temperature.

NEUTRON FLUX DISTRIBUTIONS

Experimental measurements of the neutron flux distributions in a reactor core are compared with those predicted by the nuclear design model. In addition the effects of the reflector around the reactor core and the shutdown reactivity (all control rods in the core) can be determined. The neutron flux distributions also help verify calculations of heterogeneity effects in the core.

Flux distributions can be measured by a variety of materials which become radioactive upon the absorption of a neutron. They must have a sufficiently long half-life so the induced radioactivity can be counted in a nuclear counter after the irradiation. Some flux-plotting materials which are commonly used in critical assemblies are enriched uranium, manganese, dsyposium, gold, and indium. These materials are placed in the core in the form of wires and foils.

Small counters containing special nuclear materials are also used to obtain neutron flux distribution. In these counters the neutrons are absorbed by the atoms of the material to produce energetic nuclear particles. These particles are detected by the counter and recorded in an electronic counting system. The number of pulses is proportional to the number of neutrons which are present at the location of the nuclear material.

Future Role of These Techniques in Reactor Design

In spite of the obivous risks in trying to predict the future in such a new field as reactor design, the authors feel compelled to give a brief discussion in this area. First consider analytical techniques. The goal of these techniques is to reduce the amount of experimental and analytical work required in future reactor designs. The ultimate dream of the designer is to develop the nuclear model and the calculational techniques to a point where experimentation only proves the validity of the techniques and does not contribute anything new. It may be possible to accomplish sufficient progress over the next decade, for pressurized water reactors, that our analytical ability alone will be sufficient for reactor design. How can this be brought about? The digital computer and our nuclear design model will be refined so the problems of transport theory can be reduced by experience and empirical techniques to a state where they will be recognized as engineering problems.

This should not be interpreted as meaning that no experimental work will be necessary a decade from now. Obviously the basic concepts of reactors will need to be explored, just as such concepts are in the electrical industry. Today, the designers are trying to understand the phenomena involved. As they progress up the road to greater refinement, the critical experiments, the engineering mock-up, for example, may well develop into something like a reactivity laboratory. In such a laboratory it will only be necessary to study design details, some unusual phase of a control rod design, or perhaps a unique method of reactivity control. In such a reactivity laboratory, designers would no longer need such things as the clean critical experiment.

Another function which may well develop in this laboratory of the future is reactivity adjustment or shimming. When reactor design involves more than just the design of a single unit, that is, more of the mass-production type of manufacture, it may be economical to adjust the reactivity of a core after it has been produced. To accomplish this, the addition of fuel or nuclear poisons in specific locations in the core designed to accommodate them, may be one of the more important activities in this reactivity laboratory.

Reactor physics experiments utilizing clean critical assemblies, irradiation facilities, electro-nuclear machines, kinetic experiments, and neutron spectroscopy will provide the fundamental knowledge required to simplify reactor design techniques. It is probable that design techniques unknown today will evolve from the understanding gained from the research in the foregoing fields.

With the possibility of space travel becoming more realistic, the nuclear reactor of the future may not be limited to the central power station nor to naval vessel propulsion. Nuclear reactor designs for use in space travel will obviously change the emphasis on reactor weight and will also have great impact on the biological shielding and specific power of the reactor. If this application holds promise in the future, one can expect a great deal of analytical and experimental work in the field of the interaction of radiation and matter for shielding purposes.

Glossary

Control material = movable neutron absorbing material inserted in reactor core to compensate for changes in reactivity during operation.

Criticality = the state of a reactor in which the power or neutron flux is independent of time, without an external source of neutrons. Here reactivity equals zero, or multiplication constant equals unity. Term also used as colloquial equivalent of reactivity.

Cross section = a property of nuclei representing the probability that, under given conditions, certain reactions will take place. For neutron interactions, the cross sections may be for absorption, fission or scattering.

Moderator = a substance having a large scattering cross section and small absorption cross section for neutrons, included in a reactor to slow down neutrons to energies where fission readily occurs.

Poison = a material highly absorbent of neutrons. May be included for control purposes, or may be present in structure, moderator, and fission products.

Reactivity = a quantity determining the rate at which the power level of a reactor is changing with time. Equal to $(k_{eff}-1)/k_{eff}$, where k_{eff} is the effective multiplication constant of the core.

Reflector = a moderator which surrounds the core to reflect neutrons back which would have escaped. This process results in a core of given geometrical size being more reactive with a given fuel loading than the same reactor without a reflector.

Resonance absorption = absorption of neutrons in nuclei at certain energies, where the absorption cross section is very large (resonances).

Shutdown = term used to express the reactivity or the effective multiplication constant of a core when all control elements are fully inserted.

Xenon transient = the time variation in the effective cross section and core reactivity, following power operation and shutdown, due to the highly absorbing fission product xenon. The quantity of xenon in the core builds up during operation to a steady value proportional to neutron flux and after shutdown builds up from the radioactive decay of the fission product iodine while decaying radioactively itself. The amount of xenon thus reaches a maximum some time after shutdown and then decreases again.

References

1. Engineering Description of a Water Moderated Flexible Critical Facility Using Metal Fuel, E. S. Lemborsky. *AIEE CP 58-1316* (available on request).

2. A Survey of Numerical Methods in a Solution of Diffusion Problems, G. Bilodeau, L. Hageman. *WAPD-TM-64*, Westinghouse Atomic Power Division, Pittsburgh, Pa., July 1957.

3. Synthesis of Three-Dimensional Power Distributions for a Non-Uniformly Depleted Core, R. R. Schiff. *Bettis Technical Review, WAPD-BT-4*, vol. 1, no. 4, Oct. 1957.

4. Candle—A One-Dimensional Few Group Depletion Code for the IBM-704, O. J. Marlowe, P. A. Ombrellaro. *WAPD-TM-53*, May 1957.

5. Neutron Slowing Down in Group Diffusion Theory, R. L. Hellens. *WAPD-114*, May 1956.

A-C Shocks of Varying Parameters Affecting the Heart

W. B. KOUWENHOVEN
FELLOW AIEE

G. G. KNICKERBOCKER
ASSOCIATE MEMBER AIEE

R. W. CHESNUT
NONMEMBER AIEE

W. R. MILNOR
NONMEMBER AIEE

D. J. SASS
NONMEMBER AIEE

IT IS WELL KNOWN that both the human and the dog's heart may be thrown into ventricular fibrillation by 60-cycle a-c shocks, resulting in the cessation of circulation followed by death. There are, however, little data available as to the magnitude of the current of various durations that will result in fibrillation, nor is there much information as to how the induction of fibrillation is influenced by the timing of the shock in the heart cycle. Ferris, Williams, Spence, and King[1] found that the sheep's heart was most vulnerable to shocks that fell during the *T* phase of the electrocardiogram of the heart.

The effect on the dog's heart of 60-cycle a-c shocks of 1/120, 1/60, 1/12, 1/6, 1/3, 1, 2, and 5-second durations was investigated for the current path from the right foreleg to the left hind leg. At each shock duration the minimum current needed to throw the ventricles into fibrillation, and the vulnerable period, if any, were determined. When fibrillation occurred it was arrested with the a-c closed-chest defibrillator, and normal sinus rhythm was restored.

Objectives

The objective of this study was twofold:

1. To investigate the effects of a-c shocks of varying durations, occurring at different places in the heart cycle, and to determine the magnitude of the minimum current and its position in the heart cycle that would result in fibrillating the heart.
2. To provide an opportunity for a thorough test of the a-c closed-chest defibrillator.[2,3]

The dog was used as the experimental animal in this investigation.

Heart Cycle

In order to identify the various phases of the heart cycle, a typical electrocardiogram (ECG) and its relationship to activity in the heart are shown in Fig. 1. A current flow throughout the heart is associated with its action, and this current flow produces a potential of a few millivolts at the extremities which is recorded as the ECG. The components of the ECG are the *P* wave, the QRS complex, and the *T* wave.

With the heart polarized, the action begins with the *P* wave, which originates in the sino-auricular (SA) node and which precedes the contraction of the auricles. The excitation initiated by the *P* wave travels to the auricular ventricular (AV) node and from there to the right and left ventricles, producing the *QRS* complex, which takes place just prior to the contraction of the ventricles. Following the contraction, the heart is in a depolarized state for a brief period. Then the *T* wave appears, which represents the repolarization phenomena. The repolarized heart rests for several milliseconds before the beginning of its next cycle.

Equipment

This investigation covering shocks of various durations and magnitudes required the utilization of special and diversified equipment, in the following four categories: 1. Recording equipment for measuring physiological and electrical variables. 2. Sources of adjustable voltages. 3. Provision for controlling the instant in time when the shock was initiated. 4. Means of controlling the length of the shock.

Recording Equipment

The recording equipment was a 4-channel Sandborn Poly Viso Recorder. This made it possible to record not only the ECG, but also the shocking current and the blood pressure simultaneously, and accurately to determine the instant in the heart cycle when the shock was applied.

1. *Physiological Variables*

Since the electric shocks were sent through the animals from the right foreleg to the left hind leg, the ECG potentials for recording the heart action at first were picked up by leads attached to the other

---

Paper 59-14, recommended by the AIEE Safety Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted September 3, 1958; made available for printing November 10, 1958.

W. B. Kouwenhoven, G. G. Knickerbocker, R. W. Chesnut, and W. R. Milnor are with The Johns Hopkins University, Baltimore, Md., and D. J. Sass, formerly a graduate student at Johns Hopkins, is now with The Martin Company, Baltimore, Md.

The authors wish to acknowledge and to thank the National Heart Institute, U. S. Public Health Service, for a grant in support of this study.

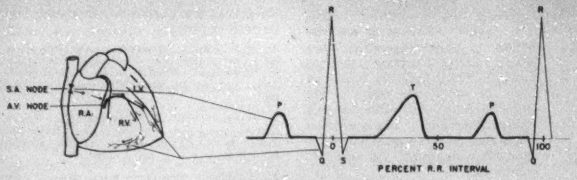


Fig. 1. The heart and its electrocardiogram

R.A. = right auricle
L.V. = left ventricle
R.V. = right ventricle

extremities. The high voltages used in applying the shock interfered with the ECG traces and made reliable records difficult to obtain. It was found that satisfactory records were regularly obtained from a small electrode on the chest over the heart and an electrode on the right hind leg.

The blood pressure was recorded on the second channel of the recorder by means of a Statham Strain Gauge Pressure Transducer connected to a cannula in either a femoral or carotid artery.

2. *Electrical Variables*

A small known noninductive resistance was inserted in series with the grounded end of the high-tension side of the shocking transformer. The voltage drop across this resistance produced by each shock was recorded on the third channel. Thus the instrument simultaneously recorded, not only the heart actions and the blood pressure, but also the magnitude, duration, and position in the heart cycle of the current. In a number of runs a potentiometer was connected across the shocking circuit and a small fraction of the high voltage was recorded on the fourth channel of the recorder. This served as a check on the duration and position of the shock.

For the 1/120- and 1/60-second shocks the response of the recorder was not sufficiently good to register these peak currents. Therefore, in order to obtain magnitude of these currents it became necessary to photograph their current traces on a calibrated oscilloscope. The recorder, however, served to indicate their position in the heart cycle.

VOLTAGE SUPPLY

In all of the experiments transformers were used to isolate the animal circuit from the primary a-c power supply. Two 5-kva transformers of different ratios were used, one connected either 1 to 2.5 or 1 to 5, and the other either 1 to 10 or 1 to 20. The fine control of the shocking potentials, essential to the determination of the minimum fibrillating current at a given shock duration, was provided by a 40-ampere continuous-duty Variac autotransformer inserted in the primary circuit of the transformer.

SHOCK POSITION CONTROL

In order to obtain a signal to position the shock at the desired point in the heart cycle, a square-wave generator with a variable-delay circuit was used. This generator was synchronized with the *R* wave of the ECG, and the position of its square-wave output with respect to the heart cycle could be set at any point. The synchronizing voltage conveniently enough came from an *R-C* circuit, which shaped the saw-toothed voltage of the horizontal deflection circuit of the scope to a sharp spike in phase with the start of the sweep and the upstroke of the *R* wave. Thus the output of the square-wave generator could be positioned at any desired point in the heart cycle.

The brief square-wave pulse was fed to the input of the unused channel of a dual-beam oscilloscope where its relative position in the heart cycle could be observed at all times. When it was desired to apply the shock to the animal a button was pressed which applied the generator pulse to a small fast-acting relay. This relay served to actuate the timer equipment which regulated the duration of the shock.

DURATION OF SHOCK

Two different methods were used to regulate the duration of the shocks. For shocks ranging from 1/120 (1/2 cycle of 60 cycles) second to 1/6 second (10 cycles), a special circuit gated the grid circuits of two back-to-back thyratrons in the primary circuit of the supply (the higher voltage, either supply) transformer. The gating circuit was adjusted to count

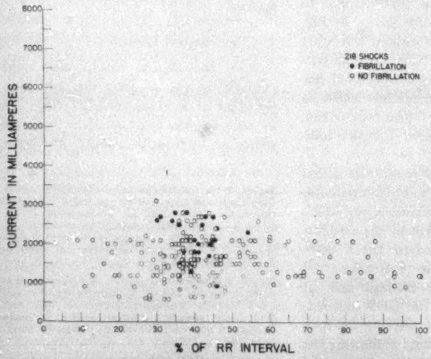


Fig. 2. 1/120-second a-c 60-cycle shock

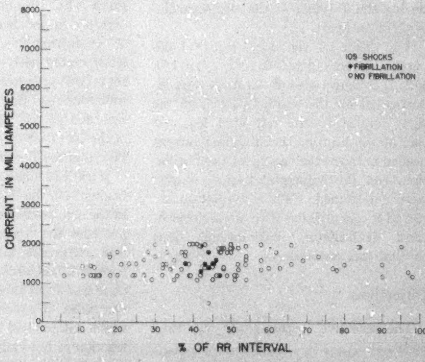


Fig. 3. 1/60-second a-c 60-cycle shock

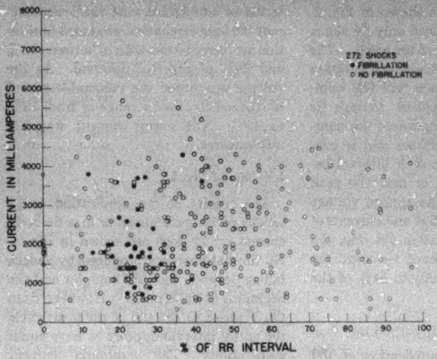


Fig. 4. 1/12-second a-c 60-cycle shock

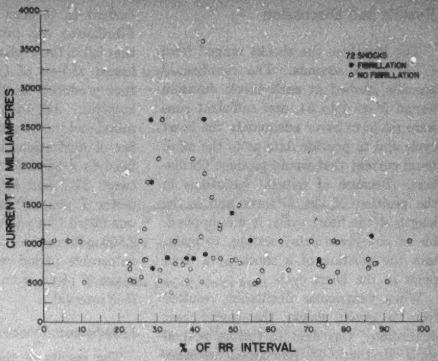


Fig. 5. 1/6-second a-c 60-cycle shock

off a certain number of cycles. It operated upon receipt of the starting signal from the quick-acting relay actuated by the square-wave generator. Thus the thyratrons acted as a precision controlled timer switch in the primary of the transformer, and the shock always started at the zero point of the a-c wave.

For shocks ranging from 1/3 to 5 seconds, the duration was regulated by an adjustable mechanical timer driven by a synchronous clock motor. The signal from the square-wave positioning generator closed a relay in the timer which applied the shock for the duration to which the timer had been set.

Procedure

The dogs, ranging in weight from 8 to 16 kilograms, were anesthetized with morphine and ether. A trachea tube was inserted and connected to a motor-driven respirator which ventilated the lungs with a mixture of air and ether. The hair was clipped from the chest and from the extremities. A cannula was inserted in an artery and connected to the blood-pressure gage transducer. Electrodes, each 1 inch by 2 inches, were strapped on the extremities. Electrode jelly was used to insure good electrical contact.

A normal record of the heart action and the blood pressure was taken on the recorder. Then the electrocardiograph leads were disconnected from the right foreleg and left hind leg, and connected either to the other extremities or to a chest electrode and the right hind leg. The high-tension leads from the transformer were then connected to the right foreleg and the left hind leg.

The electric circuits and controls were adjusted to provide a shock of the desired voltage, duration, and position in the heart cycle. The depth of anesthesia was carefully adjusted until the eye reflex was only just present. Then the recorder was started and the actuating button pressed sending the shock through the animal. From the record, the position of the shock, its duration, the magnitude of the current, and the effect on the heart and blood pressure were noted. If fibrillation occurred, the heart was defibrillated, using the closed-chest a-c defibrillator.

If the shock resulted in fibrillation, the shocking voltage was lowered by means of the Variac, and the run repeated after the heart had returned to normal sinus rhythm. If the shock did not fibrillate the heart, the voltage was raised before the next shock.

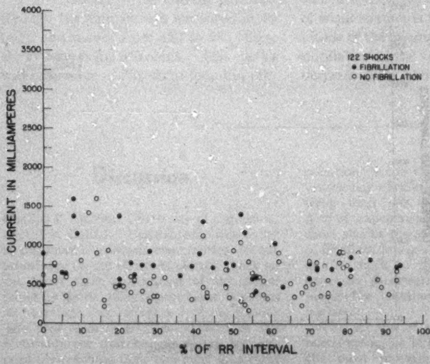


Fig. 6. 1/3-second a-c 60-cycle shock

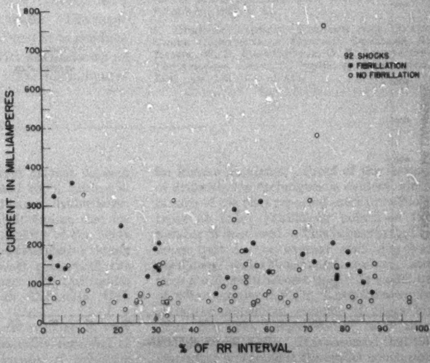


Fig. 7. One-second a-c 60-cycle shock

Results and Discussion

The length of the shocks ranged from 1/120 to 5 seconds. The number of animals studied at each shock duration varied from 4 to 34, and sufficient runs were made to cover adequately the heart cycle and to provide data as to the minimum current that would produce fibrillation. Because of natural variations in the position of the *T* wave and in the length of the heart cycle, it was impossible at any given delay setting, to maintain the position of a shock at a fixed point in the heart cycle.

When ventricular fibrillation resulted from a given shock, the closed-chest defibrillator was used to restore normal sinus rhythm. It functioned satisfactorily in every instance.

The results are plotted on eight charts, Figs. 2 through 9. The abscissa represents a complete heart cycle where the *R–R* interval is expressed as 100%. The ordinate is the current in milliamperes. The position of the beginning of each shock in the heart cycle, as well as the current, is shown by a small circle. If the shock produced ventricular fibrillation the circle appears as a black dot.

A normal dog's heart beats at a slightly faster rate than a human heart, but in these experiments the anesthetized animal's pulse ranged from 150 to 240 beats per minute, averaging about 200 beats per minute. Therefore, the average *R–R* interval is 0.3 second in these animals. In man the pulse varies from 70 to 90 beats per minute with an *R–R* interval of about 0.75 second.

1/120-Second Shock

The results of the application of 218 shocks of 1/2-cycle duration of a 60-cycle current on 14 dogs are shown in Fig. 2. Fibrillation was produced only by shock that fell in the *T* phase of the heart. The lowest current of 1/120-second duration that produced fibrillation was 925 milliamperes. As the applied voltage is raised and the current increased the number of fibrillations increases and is confined to a band in the *R–R* interval between 37% and 48%, or near the peak of the *T* phase. As the applied voltage was raised to a value that sent currents of 2,500 milliamperes through the dog, the vulnerable period widened to cover the entire *T* phase from 30% to 54% of the *R–R* interval.

1/60-Second Shock

The results of the application of 109 shocks of 1-cycle duration of a 60-cycle current on 7 dogs are shown in Fig. 3. Again the vulnerable period of the heart was found to be near the peak of the *T* phase, namely between the 38% and 47% limits of the *R–R* interval. The minimum current causing fibrillation was 1,200 milliamperes. In this test the voltage was not increased to high values as there were sufficient data to determine the minimum fibrillating shock and its location in the heart cycle.

1/12-Second Shock

The results of the application of 272 shocks of 5-cycle duration of a 60-cycle current on 34 dogs are shown in Fig. 4. Ventricular fibrillation was produced mainly by shocks falling earlier in the heart cycle than in the previous two short-shock durations. Here the vulnerable period ranged between 13% and 42% of the *R–R* interval. It should be kept in mind that a shock of this duration covers approximately 1/4 of the dog's *R–R* interval of 0.3 second. At the lower current the four shocks that produced fibrillation are confined to a span of between 22% and 28% of the *R–R* interval. As the voltage was raised the vulnerable period broadened to cover a range of from 13% to 42%. The lowest current was 720 milliamperes.

1/6-Second Shock

The results of the application of 72 shocks of 10-cycle duration of a 60-cycle current on 12 dogs are shown in Fig. 5. No fibrillations were produced by shocks falling between the *QRS* complex and the beginning of the *T* phase. However, from the *T* phase to the end of the heart cycle the entire interval appears to have equal vulnerability. The minimum current producing fibrillation was 690 milliamperes. The heart is vulnerable from the start of the *T* wave (28%) to near the end (87%) of the heart cycle.

1/3-Second Shock

The results of 122 shocks of 20-cycle duration of a 60-cycle current applied to 21 dogs are given in Fig. 6. The duration of these shocks covered the entire cycle of the dog's heart and the results show clearly that the vulnerable period extends over the entire heart cycle. The minimum current producing fibrillation was 347 milliamperes.

One-Second Shock

The results of 92 1-second shocks of 60-cycle current applied to 15 dogs are shown in Fig. 7. The results show clearly that the vulnerable period extends over the entire heart cycle when the duration of the shock covers more than one cycle. The minimum current required to induce fibrillation was 71 milliamperes.

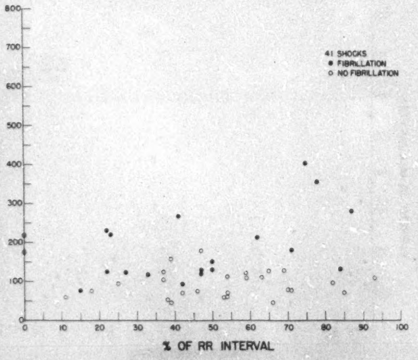


Fig. 8. Two-second a-c 60-cycle shock

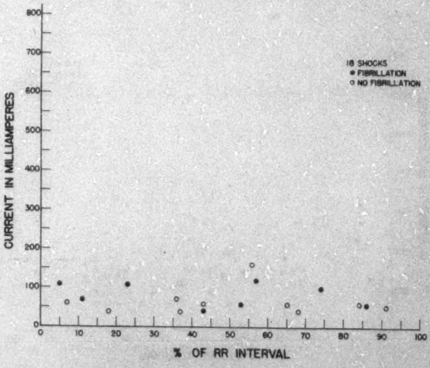


Fig. 9. Five-second a-c 60-cycle shock

Table I. Magnitude of Threshold Current

Shock Duration, Seconds	Threshold Current			
	Current, Milliamperes			
	Lowest	Second	Third	Average
1/120	925	1,175	1,300	1,130
1/60	1,200	1,200	1,299	1,230
1/12	720	750	880	780
1/6	690	750	800	770
1/3	347	375	412	380
1	71	73	80	75
2	77	91	120	96
5	42	56	58	52

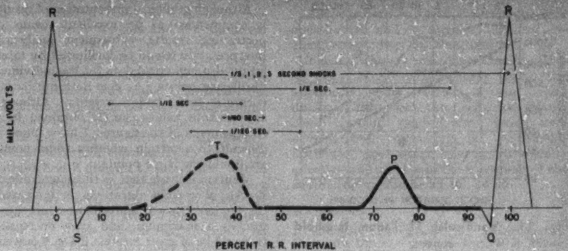


Fig. 10. Vulnerable period, dog's heart

Two-Second Shock

The results of 41 2-second 60-cycle shocks applied to 10 dogs are shown in Fig. 8. The results again illustrate the fact that the entire heart cycle is vulnerable when the shock duration covers several heart cycles. The minimum fibrillating current measured was 77 milliamperes.

Five-Second Shock

Only 18 5-second shocks were applied to a total of 4 animals. The results are shown in Fig. 9 and it is clear that the entire heart cycle is vulnerable. The lowest fibrillating current found was 42 milliamperes.

Threshold Current

We have arbitrarily selected, as a measure of the magnitude of the threshold current, the average of the three lowest currents causing ventricular fibrillation. These are given in Table I.

It is evident that as the duration of the shock increases the fibrillating current decreases. Contacts of several seconds duration with 50-volt a-c circuits may result in a fatal accident.

RESISTANCE

The resistance of the current pathway through the animals was measured in 40 tests, and ranged from 420 to 930 ohms, the average being 650 ohms. This checks with the value of 700 ohms found in previous tests[4] using this same current pathway.

VULNERABLE PERIOD

Fig. 10 presents the spans of the heart cycle covered by the beginnings of the applied shocks of different durations that produced fibrillations. These spans are superimposed on an ECG to show the vulnerable periods found. The *T* wave, which usually varies slightly in position and length, is shown as a broken line covering its normal range of variation.

As the duration of the shock decreases the vulnerable period of the heart cycle narrows and for very short shocks is located in the *T* phase.

Summary

This paper presents the results of an investigation of the production of ventricular fibrillation caused by shocks ranging in duration from 1/120 second to 5 seconds. A vulnerable period in the vicinity of the *T* phase of the heart cycle exists for short shocks. The entire heart cycle is vulnerable to shocks that are of sufficient duration to cover it, irrespective of when contact is first made. The magnitude of the minimum current to produce fibrillation of the ventricles was shown to decrease from about 1,200 milliamperes for 1/120-second shocks to near 50 milliamperes for 5-second shocks for a current pathway from right foreleg to left hind leg in an ether-anesthetized experimental animal. The resistance of this pathway averaged 650 ohms. When fibrillation occurred it was successfully arrested with the a-c closed-chest defibrillator.

Since it is known[5] that as the magnitude of shocks of certain durations is increased the tendency to cause fibrillation decreases, it would be of interest to extend this study to determine whether this holds true for shocks of all durations.

References

1. EFFECT OF ELECTRIC SHOCK ON THE HEART. L. P. Ferris, B. G. King, P. W. Spence, H. B. Williams. *AIEE Transactions*, vol. 55, May 1936, pp. 498-515.

2. FIELD TREATMENT OF ELECTRIC SHOCK CASES. W. B. Kouwenhoven, W. R. Milnor. *Ibid.*, vol. 76, pt. III, Apr. 1957, pp. 82-87.

3. CLOSED CHEST DEFIBRILLATION OF THE HEART. W. B. Kouwenhoven, W. R. Milnor, G. G. Knickerbocker, W. R. Chesnut. *Surgery*, St. Louis, Mo., vol. 42, 1957, p. 550.

4. THE EFFECTS OF HIGH VOLTAGE, LOW CAPACITANCE ELECTRICAL DISCHARGES IN THE DOG. W. B. Kouwenhoven, W. R. Milnor. *Transactions*, Professional Group on Medical Electronics, Institute of Radio Engineers, New York, N. Y., vol. PGME-11, 1958, p. 41.

5. CURRENT FLOWING THROUGH THE HEART UNDER CONDITIONS OF ELECTRIC SHOCK. D. R. Hooker, W. B. Kouwenhoven, O. R. Langworthy. *The American Journal of Physiology*, Washington, D. C., vol. 103, 1933, p. 444.

◆

Discussion

Charles F. Dalziel (University of California, Berkeley, Calif.): Undoubtedly this paper is the most significant contribution to our knowledge on the effects of electric shock on the heart since publication of the AIEE paper by Ferris, King, Spence, and Williams 23 years ago (see reference 1 of the paper). The profession is greatly indebted to Dr. Kouwenhoven and his associates for successfully reaching their two most important objectives, namely the quantitative determination of the minimum current causing ventricular fibrillation and 100% success in using their a-c closed-chest defibrillator. Also of importance is their finding that for short shocks the susceptibility of the heart to fibrillate depends upon the relative timing of the shock to the heart cycle, and the generally good agreement with the results previously obtained at Columbia University.

The success of the defibrillator is very encouraging, as it is generally regarded as the most promising means of arresting ventricular fibrillation thus far developed for human accidents. Proof of the success of defibrillation techniques is evident, since in each of the tests reported there are almost twice as many fibrillating points as the number of dogs used. This is interpreted to mean that on the average each dog was fibrillated, defibrillated and revived, and then fibrillated a second time. It would be of great interest for Dr. Kouwenhoven to state the maximum number of times his prize dog was "electrocuted" and then brought back to life. It is assumed that the statement will confirm the earlier conclusion by Ferris and his associates that if the heart

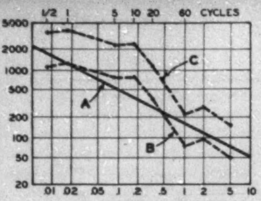


Fig. 11. Ventricular fibrillation threshold currents

A—Current that will not cause ventricular fibrillation in men determined by Dalziel in reference 1 of this discussion

B—Ventricular fibrillation threshold for dogs as determined in the paper

C—Threshold for dogs multiplied by 3 to approximate threshold for men

is not damaged and is given time to return to normal, successive shocks have no cumulative effect on the susceptibility of the heart to fibrillate.

The vulnerable period of the dog's heart illustrated in Fig. 10 is most interesting. It is entirely possible that the period when the heart is susceptible to fibrillate may span even shorter and shorter percentages of the *T* phase as the shock duration is further decreased. This may explain why accidental deaths due to impulse shocks, such as those associated with lightning generators and capacitors, are so infrequent.

The following ideas are suggested with the thought of improving the paper. The authors state that for shocks ranging from 1/120 to 1/6 second the shocks always started at the zero point of the a-c wave. This statement is interpreted to mean that the shock current was sinusoidal, and therefore the d-c component was zero. For shocks ranging from 1/3 to 5 seconds the duration was regulated by a mechanical timer and it is likely that the shocks occurred at random with reference to the shock-voltage wave. However, since the current was limited by variable-ratio transformers in series with the body resistance of the animal which averaged some 650 ohms, the *X-R* ratio was probably small and the resulting d-c component of shock current, even when present, would likely be of relatively short duration with respect to the duration of the shock. For these reasons it is likely that the fibrillating currents given in Table I and in Figs. 2–6 were nearly symmetrical rms milliamperes. In the interest of both clarity of presentation and possible subsequent application to accidents, it is suggested that the ordinates of Figs. 2 to 9 be labeled "Current in Symmetrical Rms Milliamperes."

It is requested that the average body weight of the animals fibrillated be inserted directly on Figs. 2 to 9 or given in tables, or in the text in paragraphs discussing the separate shock-duration tests. It would seem that the average weight should be computed from body weights for each fibrillating point, that is, animals fibrillated more than once should be considered the same as if they were new animals.

Although graphical presentation of results is commended as an excellent means of portraying results, subsequent analytical interpretation would be facilitated by tabulating also numerical data for the individual fibrillating points, i.e., give both fibrillating currents and body weights. Although approximate values can be obtained from enlargements of the figures, I have found it difficult to ascertain whether some points are double or triple points.

Mention is made that as the magnitude of shocks of certain durations is increased, the tendency to cause fibrillation increases, reaches a maximum, and then decreases. It would be of great interest if Dr. Kouwenhoven and his associates would find it possible to extend the study to determine whether this holds for shocks of all durations.

As an experimenter in this same general field, I feel very special recognition and sincere thanks are due these investigators and their sponsors for the great expenditure of time, thought, and resources, and the meticulous care required to observe, analyze, and record the 944 shock tests reported in this paper.

R. X. French (Sargent & Lundy, Chicago, Ill.): This paper is a valuable addition to the knowledge of the effect of electric shocks on heart action. Of major interest is the fact that the experimental results agree with the threshold fibrillation current of 100 milliamperes for shock duration in the order of 1 to 3 seconds determined by L. P. Ferris, et al, in 1936 (reference 1 of the paper). The most valuable contribution of this paper is the experimental verification of C. F. Dalziel's analytical derivation of the relationship between shock duration and fibrillation threshold current.[1]

The solid line on Fig. 11 shows Dalziel's threshold currents plotted against shock duration. It represents the current that $99^1/_2\%$ of all men could withstand without ventricular fibrillation. Curve *B* is plotted for the average of the three lowest currents that produced ventricular fibrillation in the dogs used in the experiments covered by the paper. Curve *C* is plotted for the dog threshold currents multiplied by 3 in order to approximate the threshold current for men. The multiplier is somewhat arbitrary. It was determined from the relative body weights of the dogs used in the experiments and data presented in reference 1 of the paper.

Comparison of the approximate threshold for men with Dalziel's curve (*A*) shows a reasonable safety factor and verifies that Dalziel's criterion for safe current should be used in the shock-time range of 0.01 second to 5 seconds. This criterion was used in the recent AIEE Committee Report on surface gradients.[2]

It would be interesting to have the authors discuss possible reasons for the dip in the threshold current at 1-second shock duration. Also, any comments that they may have on the apparent flattening of the curve at the low-shock duration end would be of interest since it is usually considered that the safe current increases with shorter shock durations, particularly for shocks of the impulse variety. Do the authors believe that the anesthetic affected the dog's response to shocks? Do the authors have any comments on converting animal data to data for men?

REFERENCES

1. A STUDY OF THE HAZARDS OF IMPULSE CURRENTS, **Charles F. Dalziel.** *AIEE Transactions*, vol. 72, pt. III, Oct. 1953, pp. 1032–43.

2. VOLTAGE GRADIENTS THROUGH THE GROUND UNDER FAULT CONDITIONS, **AIEE Committee Report.** *Ibid.*, vol. 77, pt. III, Oct. 1958, pp. 669–92.

W. B. Kouwenhoven, G. G. Knickerbocker, R. W. Chesnut, W. R. Milnor, D. G. Sass: The authors are not able to name their "prize dog" as defined by Professor Dalziel. However, one animal after having been thrown into ventricular fibrillation nine times, and defibrillated each time, was retired to his cage for two months. At the end of this period he was sacrificed and subjected to autopsy. No evidence of trauma was found.

It is possible, as suggested by Professor Dalziel, that the period of the dog's heart when it is most susceptible to fibrillation, may span even shorter percentages as the shock duration is decreased. We made a total of 129 experiments on 21 mongrel dogs to determine vulnerability to single stimuli consisting of capacitor discharges through

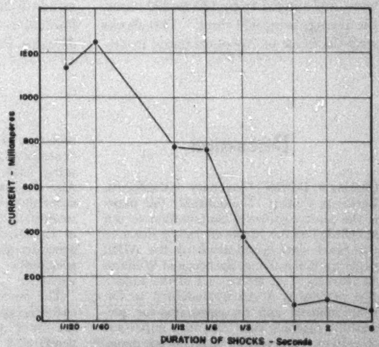


Fig. 12. Average of lowest three fibrillating currents versus shock duration

external electrodes on the thorax.[1] Capacitors of 25 to 50 microfarads charged to voltages of 1 to 4 kv were used. Voltages from 2 to 4 kv produced multiple ventricular extra systoles usually followed by a gradual return to normal rhythm. In 9 out of 35 experiments with peak voltages of 2.7 to 4 kv the ventricular extra systoles developed into an established ventricular fibrillation. Susceptibility to these effects extended throughout most of the heart cycle and was not limited to the relatively short late systolic period.

The shocking currents in our study were supplied by the local power system and may be considered as 60-cycle rms currents with a zero or negligible d-c component.

The weights of our animals were recorded but we failed to find any correlation between the body weights and the minimum fibrillating current. The objective of our study was to determine the effects of shocks applied at different points in the heart cycle, and no attempt was made to ascertain the minimum fibrillating current of each animal at each point in the heart cycle.

The discussion submitted by Mr. French sets forth certain conclusions to which the authors find it difficult to subscribe. When the minimum fibrillating currents are plotted, as shown in Fig. 12, it appears that there exist at least two, if not three, plateaus. It is our opinion that the effect on the heart of electric shocks of varying durations depends upon more than a single biological mechanism or reaction.

All dogs used in these experiments were anesthetized and we have no data for unanesthetized dogs for comparison.

The authors are not able either to affirm or deny Mr. French's query as to the application of these results to man. They do know that both the open- and closed-chest electrical defibrillators, which were developed for arresting the ventricular fibrillating dog's heart, have proved equally effective and reliable in the treatment of the fibrillating human heart.

REFERENCE

1. CARDIAC RESPONSES TO TRANSTHORACIC CAPACITOR DISCHARGES IN THE DOG, W. R. Milnor, G. G. Knickerbocker, W. B. Kouwenhoven. *Circulation Research*, New York, N. Y., vol. 6, 1958, p. 60.

Eddy-Current Losses in Solid and Laminated Iron

PAUL D. AGARWAL
ASSOCIATE MEMBER AIEE

Synopsis: In this paper formulas are developed, on the basis of simple physical conceptions, for the calculation of eddy currents in solid and laminated iron, which, as the experiments presented show, enable the losses caused by eddy currents to be predicted with surprising accuracy. A simple formula for the calculation of the power factor (pf) that the magnetic circuit reflects into the magnetizing winding is also presented.

Solution of Maxwell's equations, when a sinusoidal mmf (magnemotive force) is impressed on a core composed of steel plates, and under the conditions of an ideal rectangular B-H curve, Fig. 1, gives a wave of flux density B_0 that enters each plate from the outside surface at the start of each half-cycle, and penetrates to a depth δ, the "depth of penetration." If δ is less than d, the half thickness of the plate, the inner part of the plate remains unmagnetized, and the magnetic flux as well as the eddy currents are confined to a layer of depth δ on each surface of the plate. If δ is greater than d, the flux-density waves from the two sides meet in the center of the plate before the end of the half-cycle, and for the remainder of the half-cycle the plate remains magnetized to the value B_0. In this case, therefore, the eddy currents flow during the first portion of each half-cycle and are zero during the period the plate is fully magnetized. Waveshapes of voltages induced in search coils placed around the core substantiate this experimentally. The eddy-current loss and the pf of the magnetic circuit are calculated by the Poynting vector theorem.

Since the assumption of a rectangular B-H curve is not exact, it is necessary to choose a limiting value of the saturation flux density that will provide the best agreement between theory and test. After some study, the best value of B_0 to be used in the formulas has been determined empirically to be 3/4 B_m, where B_m is the flux density obtained from the static magnetization curve of the material corresponding to H_m, the peak of the impressed mmf. With this value of B_0, the formula gives the eddy-current losses with satisfactory accuracy over the entire range of tests performed.

Tests were made over a wide range of variables: plate thickness 0.025–1 inch; frequency from 15–2,000 cps (cycles per second); and applied mmf from 25–450 peak-ampere turns per inch. The calculated and measured values of eddy-current losses are within about 15% of each other over the entire range of the tests. In different samples of the same material, the value of H for the same B may vary within ±10–15%, and hence an error of this order is to be expected in estimating these losses.

WHEN IRON is subjected to an alternating mmf, energy is lost in the iron owing to the phenomena of eddy currents and hysteresis. In most applications the designer is interested in minimizing these losses to improve the efficiency The development of magnetic materials and techniques has kept pace with this requirement. However, many parts of machines cannot be adequately laminated, and solid surfaces are exposed to alternating fields; such as, end plates of rotors and stators in the field of the end turns, transformer tanks in the field of current leads, and numerous others. On the other hand, there are applications like magnetic clutches and eddy-current brakes, where one tries to maximize these losses.

Therefore, the problem of estimating losses in solid and in laminated iron has been of prime concern to the research scientist and the design engineer. Magnetic circuits of transformers and rotating machines are built up of thin laminations and this leads many to think that with a fixed applied mmf the losses will increase with the thickness of laminations and would become excessive if the limit is reached; i.e., if the laminated body is replaced by solid steel. Because the shielding effect of eddy currents may actually reduce the losses, this concept is wrong. For a given total flux, however, the losses will increase steadily with increasing lamination thickness.

In spite of the enormous amount of work done in this field in the past 50 years, no theoretically sound formula has been developed which would give the iron losses within engineering accuracy over a wide range of material thickness, applied mmf

---

Paper 59-12, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted September 16, 1958; made available for printing November 10, 1958.

PAUL D. AGARWAL is with the University of Massachusetts, Amherst, Mass., and Consultant to the General Electric Company. This material is part of the thesis submitted for the degree of Doctor of Electrical Engineering at the Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

The author is greatly indebted to the General Electric Company for its support of this research, and to Philip L. Alger, Consulting Engineer of that company, for his inspiring guidance and assistance, without which this work would not have been possible. The author would also like to express his deep appreciation to Dr. Ernst Weber, President of the Polytechnic Institute of Brooklyn, and Chairman of his Guidance Committee, for his critical review, helpful suggestions, and kind interest.

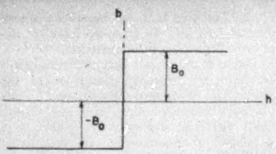


Fig. 1 (left). Assumed magnetization curve

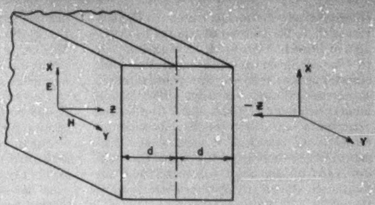


Fig. 2 (right). Plane waves penetrating inward from both faces of plate of thickness 2d

and frequency. The difficulty in deriving such a formula arises from the extremely nonlinear relation between B and H, which the hysteresis loop represents for a ferromagnetic material. Therefore, an exact mathematical calculation of the iron losses is not possible.

An approximate solution obtained by assuming the permeability to be a constant may yield good results so long as the material is being worked on the linear portion of the magnetization curve. However, it should be realized that even for weak magnetization a close examination of the magnetization curve will show that the assumption of linearity in actual practical cases is not justified. Also, at low densities hysteresis loss is predominant, and the division of the total measured losses into eddy-current losses and hysteresis losses depends on the assumptions made. The accuracy of these assumptions and, hence, the division of experimental losses is questionable. Since in most applications iron is saturated, the permeability will vary over a wide range $\mu_{\max}/\mu_{\min}=10\text{–}500$, more or less. Since the eddy-current losses as given by the linear theory are dependent upon the value of permeability, the results given by the linear theory will have little quantitative meaning except to give a physical conception of the phenomenon.

A careful review of the literature shows that many authors[1-3] have employed the rectangular B-H curve assumption in deriving their results, and MacLean[2] has obtained in this way formulas for the solid iron case. His analysis has been extended in this paper to the case of thin plates thus yielding the general formula. However, so far as the author is aware, no previous writer has derived the formulas for the laminated iron case that are developed here. Nor have any authors obtained satisfactory checks between calculated and test results over so wide a range of applied mmf, frequency, and plate thickness as are covered here.

Analysis

The assumed ideal magnetization curve under which the eddy-current phenomena are being analyzed is illustrated in Fig. 1.

Consider an iron plate of infinite dimensions in the xy plane, with its thickness $2d$ in the Z direction as shown in Fig. 2. Let a magnetizing winding, not shown in Fig. 2, produce and maintain on each face of the plate a uniform sinusoidal magnetizing force $H=H_m \sin \omega t$, in the Y direction with no variation in the X direction. The corresponding electric field will be normal to the magnetic field and, hence, in the X direction. Thus X-polarized plane waves impinge on the outside faces of the plate and propagate inward toward the center of the plate. Since this wave penetration is symmetrical with respect to the middle of the plate, the phenomena in the half thickness of the plate only needs to be considered. The waves enter the plate from the surface and penetrate inward with resulting attenuation. The movement of the front of the wave $h=0$ is of particular interest because, as a result of the assumed B-H curve, change in magnetization takes place only at this point. The wavefront $h=0$, with h increasing, is accompanied by a flux-density wave of constant amplitude $+B_0$. Thus, flux waves of density $+B_0$ penetrate from the faces into the plate. The extreme depth to which fields penetrate and beyond which there are no fields at all is called the "depth of penetration" δ. It should be noted that the term "depth of penetration" in the case of linear permeability has a different connotation.

Two cases now arise, one in which the depth of penetration δ is less than d, the half thickness of the plate, and the other in which δ is greater than d. In the first case, the fields penetrate to depth δ on each side of the plate and the inner part of the plate remains unmagnetized. In the second case, the two flux waves will meet at the center of the plate before the end of the half-period, and for the remaining part of the half-period there will be no change of flux in the plate. This induction remains in the form of remanent flux until it is annihilated by a negative field force.

At the end of the half-period, h at the surface becomes zero and is at the beginning of its negative half-wave. Now again a wavefront $h=0$, but with h decreasing, penetrates into the plate. This wavefront is accompanied by a flux-density wave of constant amplitude $-B_0$, and reverses the direction of magnetization in the plate. Thus, the plate is alternately magnetized to density $\pm B_0$ up to the depth δ or d whichever is smaller. The transition of magnetic induction takes place at a point $\zeta(t)$, which defines an xy plane, referred to here as the "separating surface." The separating surface moves parallel to itself with time.

Maxwell's Equations

Maxwell's field equations in media at rest and in integral form, neglecting displacement current, are

$$\oint \mathbf{h}\cdot d\mathbf{l} = \gamma \int_s \mathbf{e}\cdot d\mathbf{s} \tag{1A}$$

$$\oint \mathbf{e}\cdot d\mathbf{l} = \frac{\partial}{\partial t}\int_s \mathbf{b}\cdot d\mathbf{s} \tag{1B}$$

where γ is the conductivity and all variables are expressed in meter-kilogram-second units.

In the problem under consideration, the electric field has only an x-component e, and the magnetizing force and magnetic field only y-components h and b

$\mathbf{e}=\mathbf{i}e$

$\mathbf{h}=\mathbf{j}h$

This reduces Maxwell's equations to the following

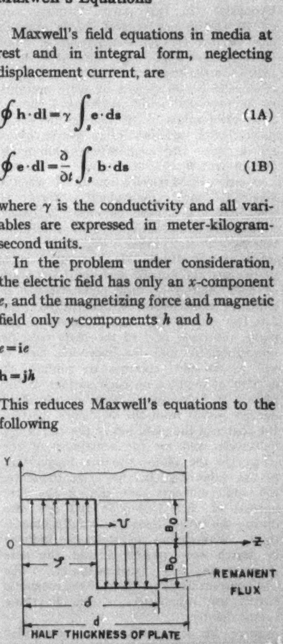


Fig. 3. Separating surface penetrating the plate

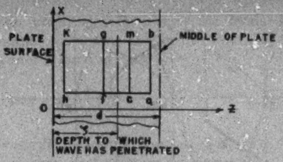


Fig. 4. XZ section of the plate

ab=cd=fg=hk
=unit length=1 meter

$$\oint h\mathbf{j}\cdot d\mathbf{l}=\gamma\int_s \mathbf{e}\cdot d\mathbf{s} \tag{2A}$$

$$\oint e\mathbf{i}\cdot d\mathbf{l}=-\frac{\partial}{\partial t}\int_s \mathbf{b}\cdot d\mathbf{s}=-\frac{\partial}{\partial t}(\phi) \tag{2B}$$

where ϕ is the flux passing through the surface around which the line integral is taken.

Let E, B, and H be the values of the variables at the surface, i.e., at $Z=0$. The solution of equations 2(A) and (B) is sought compatible with Fig. 1, the assumed magnetization curve, and with an impressed sinusoidal magnetizing force at the surface.

$$H=H_m \sin \omega t \tag{3}$$

The rectangular magnetization curve can be expressed as

$$B=B_0 \operatorname{sig} H \tag{4}$$

where sig H=sign of H and is the unit square wave associated with the sinusoidal H.

Consider a YZ section of half the plate as shown in Fig. 3. Let at any instant the separating surface be at a depth ζ and moving toward the center of the plate with a velocity v. Then in a time dt the flux $B_0\, v\, dt$ is built up and at the same time an equal amount of remanent flux is destroyed. Therefore, the rate of change of flux per unit length in the x direction is given by

$$\frac{d\phi}{dt}=2B_0 v \tag{5}$$

The line integral in equation 2(B) is easily evaluated in this case by taking the surface s in the xz plane; see Fig. 4.

Consider the closed loop $abmc$ in Fig. 4. There is no change of flux taking place in this area, which shows that there is no electric field beyond the separating surface.

Now consider the closed loop $cmgf$, which encloses the separating surface. From equations 2(B) and 5

$$e_{cm}+e_{gf}=2B_0 v$$
$$=2B_0\frac{d\zeta}{dt} \tag{6}$$

$$\text{or}\quad e_{gf}=2B_0\frac{d\zeta}{dt}\ \text{since}\ e_{cm}=0 \tag{7}$$

There is no change of flux to the left of the separating surface and, hence, considering loop $fgkh$ will show that

$$e_{fg}=e_{hk}=E \tag{8}$$

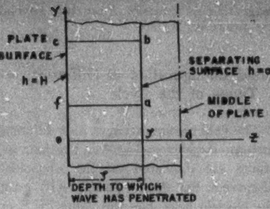


Fig. 5. YZ section of the plate

ab=cf=unit length
=1 meter

This shows that the electric field in the region to the left of the separating surface, that is, $0<Z<\delta$, is the same everywhere, and its magnitude is equal to E the value of the electric field at the outside surface, and the discontinuity in the electric field occurs at the separating surface.

To evaluate the line integral in equation 2(A), consider the loop $abcf$ in Fig. 5 with surface in the yz plane. The magnetic field intensity h is zero at the separating surface and has no component in the z direction.

$$\therefore \oint_{abcf}\mathbf{h}\cdot d\mathbf{l}=H$$
$$=\text{current enclosed in loop } abcf$$
$$=\gamma E\zeta \tag{9}$$

giving

$$\frac{\partial h}{\partial z}=\frac{H}{\zeta}=\gamma E \tag{10}$$

From the foregoing equation the following conclusions are drawn:

1. The induced eddy currents flow with uniform density in the region between the outside surface of the plate and the separating surface, and are zero beyond.
2. The magnetic field intensity h decreases linearly from its value H at the outside surface ($z=0$), to zero at the separating surface ($z=\zeta$).
3. The electric-field intensity e, and

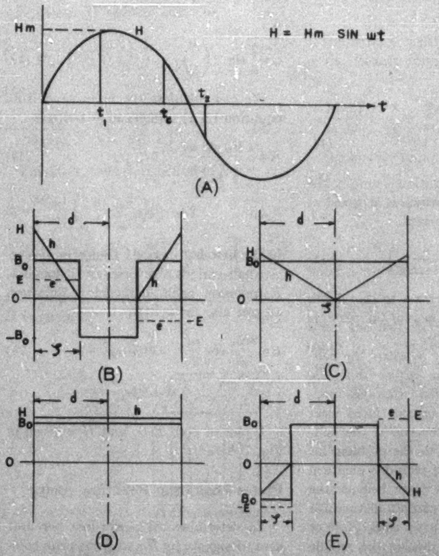


Fig. 6. Field configurations at four different instants

A—Mmf wave
B—Flux has penetrated to depth ζ. $\zeta<d$, $t=t_1$
C—Flux waves have just reached the center of the plate. $\zeta=d$, $e=0$, $t=t_0(-)$
D—After the flux wave has reached the center of the plate, h becomes uniform. $h=H$, $e=0$, $t=t_0(+)$
E—New magnetic flux wave has penetrated for a distance ζ $t=t_2$
t_0=the instant at which separating surfaces reach the middle of the plate

magnetic-field intensity h are zero beyond the separating surface.

Field Configuration at Different Times of the Cycle

The field configuration at four different instants is given in Fig. 6. The four instants are indicated on the sinusoidal H curve in Fig. 6(A). In Fig. 6(B), $\zeta<d$, the electric field, has the value $e=E$ or zero as shown, and h is linear to ζ and then zero as shown. When the separating surface moving inward reaches the center of the plate at $\zeta=d$ (say at $t=t_c$), the whole material is magnetized to saturation and no change of flux takes place for the remaining part of the half-cycle, i.e., during the interval between t_c and π/ω. After half a cycle, at $t=\pi/\omega$, the separating surface again starts from each face and reverses the direction of magnetization. During the interval t_c to π/ω, there is no change of flux in the plate and, hence, $e=0$. Substituting from Fig. 6(B) in equation 9

$$\frac{\partial h}{\partial s}=\frac{H}{\zeta}=\gamma E \qquad \zeta<d \text{ or } 0<t<t_c$$

$$=0 \qquad t_c<t<\frac{\pi}{\omega} \qquad (11)$$

This means that after t_c, h everywhere jumps to the value H at the surface, thus becoming uniform. This is shown in Fig. 6(C). After half a cycle, at $t=\pi/\omega$, the magnetic-field intensity at the surface H is equal to zero. Therefore, h is zero everywhere in the material, and a new separating surface of oppposite magnetic induction starts from the two ends. This condition is shown in Fig. 6(E).

Thus Maxwell's equations have been replaced by the following pair

$$\left.\begin{aligned}\frac{H}{\zeta}&=\gamma E\\ E&=2B_0\frac{d\zeta}{dt}\end{aligned}\right\} \qquad \begin{aligned}&0<t<t_c && (12A)\\ &\zeta<d && (12B)\end{aligned}$$

or

$$\left.\begin{aligned}h&=H\\ e&=0\end{aligned}\right\} \qquad t_c<t<\frac{\pi}{\omega}$$

Eliminating E between equations 12(A) and (B)

$$2\zeta\frac{d\zeta}{dt}=\frac{H}{\gamma B}=\frac{H_m \sin \omega t}{\gamma B_0 \operatorname{sig} H}$$

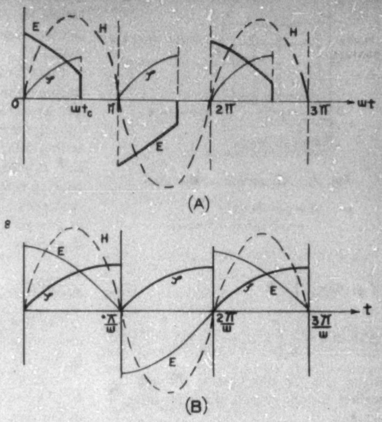
$$\frac{d}{dt}(\zeta^2)=\frac{H_m}{\gamma B_0}\left|\sin \omega t\right| \qquad 0<t<t_c \qquad (13)$$

Equation 13 shows that the transition point always moves from the outside surfaces toward the center of the plate where it ceases to exist, and repeats every half-cycle when H becomes zero at the surface.

Fig. 7. Plots of waveshape

A—H, E, and ζ

B—H, E, and ζ for the solid-iron case



Depth of Penetration

If the plate is thick enough, then the transition point ζ will penetrate to its maximum depth, depth of penetration, δ at the end of the half-cycle. The motion of the separating surface as a function of time, in the first half-period, can be evaluated as follows:

$$\frac{d\zeta^2}{dt}=\frac{H_m}{\gamma B_0}\sin \omega t \qquad \omega t<\pi$$

$$\therefore \zeta^2=\int_0^t \frac{H_m}{\gamma B_0}\sin \omega t\, dt$$

giving

$$\zeta=\left(\frac{2H_m}{\omega\gamma B_0}\right)^{1/2}\sin\frac{\omega t}{2} \quad 0<t<\frac{\pi}{\omega} \qquad (14)$$

The depth of penetration is simply the maximum value of ζ which is attained at the end of each half-period.

$$\therefore \delta=\left(\frac{2H_m}{\omega\gamma B_0}\right)^{1/2}=\left(\frac{2}{\gamma\omega B_0/H_m}\right)^{1/2} \qquad (15)$$

The fields will penetrate to this depth when d, the half-thickness of the plate, is greater than or equal to δ.

In a thin plate, that is, where the half-thickness of the plate is less than the depth of penetration, the two waves moving inward from the outside faces meet at the center before the end of the half-period. The solution of the problem for this portion of the half-period is identical with that for the thick plate, because the electromagnetic wave can not distinguish whether it is penetrating a thick plate or a thin plate, until it reaches the middle and confronts the wave approaching from the opposite side.

Therefore

$$\zeta=\left(\frac{2H_m}{\gamma\omega B_0}\right)^{1/2}\sin\frac{\omega t}{2} \qquad 0<t<t_c$$

$$=\delta \sin\frac{\omega t}{2} \qquad (16)$$

where t_c is the time when the separating surface reaches the middle of the plate.

$$d=\delta \sin\frac{\omega t_c}{2} \qquad (17)$$

or

$$t_c=\frac{2}{\omega}\sin^{-1}\left(\frac{d}{\delta}\right) \qquad (18)$$

We can calculate the value of E from equation 12(A), which can be written

$$E=\frac{1}{\gamma}\frac{H_m \sin \omega t}{\delta \sin\frac{\omega t}{2}} \qquad 0<t<t_c \qquad (19)$$

$$=0 \qquad t_c<t<\frac{\pi}{\omega}$$

in the first half-period. Since in successive half-periods H reverses and ζ does not, E reverses each half-cycle. Equation 19 can also be written

$$E=\frac{2H_m}{\gamma\delta}\cos\frac{\omega t}{2} \qquad 0<t<t_c \qquad (20)$$

$$=0 \qquad t_c<t<\frac{\pi}{\omega}$$

A plot of $\zeta(t)$, $E(t)$, and H is given in Fig. 7(A).

Power Loss from Poynting Vector

To determine the power-loss per unit area, designate by $\bar{H}$ and $\bar{E}$ the peak complex amplitudes of H and the fundamental component of E. The Poynting vector $\bar{N}$ into the metal is

$$\bar{N}=N+jM=\frac{1}{2}E\bar{H}^{*} \tag{21}$$

where H^{*} means the conjugate of H. The real part N of $\bar{N}$ gives the power flow. Since H is sinusoidal, the harmonics of E do not contribute to the power flow; see Appendix I.

A complex surface impedance for the plane wave (wave impedance at the surface) $\bar{\eta}$ is usually designated by

$$\bar{\eta}=r+jx=\frac{\bar{E}}{H} \tag{22}$$

which gives

$$N=Re\bar{N}=\frac{1}{2}rH_m{}^{2} \tag{23}$$

The Fourier coefficients for the fundamental of E are obtained using Fig. 6 and equation 20,

$$A_1=\frac{2H_m}{\gamma\delta}\frac{\omega}{2\pi}2\int_0^{t_c}\cos\frac{\omega t}{2}\cos\omega t\,dt$$

$$=\frac{2H_m}{\gamma\delta\pi}\left(\frac{1}{3}\sin\frac{3\omega t_c}{2}+\sin\frac{\omega t_c}{2}\right) \tag{24}$$

$$B_1=\frac{2H_m\omega}{\gamma\delta\pi}\int_0^{t_c}\cos\frac{\omega t}{2}\sin\omega t\,dt$$

$$=\frac{2H_m}{\gamma\delta\pi}\left[\frac{4}{3}-\frac{1}{3}\cos\frac{3\omega t_c}{2}-\cos\frac{\omega t_c}{2}\right] \tag{25}$$

Equation 17 reduces these to

$$A_1=\frac{4H_md}{\delta^2\gamma\pi}\left[1-\frac{2}{3}\left(\frac{d}{\delta}\right)^2\right] \tag{26}$$

and

$$B_1=\frac{8H_m}{3\gamma\pi\delta}\left[1-\left(1-\frac{d^2}{\delta^2}\right)^{3/2}\right] \tag{27}$$

Therefore, the fundamental of E is

$$E_1=\frac{8H_m}{3\gamma\pi\delta}\left[\left(3\frac{d}{\delta}-2\frac{d^2}{\delta^2}\right)\cos\omega t+2\left\{1-\left(1-\frac{d^2}{\delta^2}\right)^{3/2}\right\}\sin\omega t\right] \tag{28}$$

and the peak complex amplitude of E and H is given by .

$$\bar{E}=\frac{8H_m}{3\gamma\delta\pi}\left[\left(3\frac{d}{\delta}-2\frac{d^2}{\delta^2}\right)-j2\left\{1-\left(1-\frac{d^2}{\delta^2}\right)^{3/2}\right\}\right] \tag{29}$$

$$\bar{H}=-jH_m \tag{30}$$

Substituting these in equation 22, the surface impedance is

$$\bar{\eta}=r+jx=\frac{\bar{E}}{\bar{H}}=\frac{16}{3\pi\gamma\delta}\left[\left\{1-\left(1-\frac{d^2}{\delta^2}\right)^{3/2}\right\}+\frac{j}{2}\left(3\frac{d}{\delta}-2\frac{d^2}{\delta^2}\right)\right] \tag{31}$$

The phase angle θ of the complex surface impedance $\bar{\eta}$ is equal to

$$\theta=\tan^{-1}\frac{x}{r}$$

Therefore from equation 31

$$\tan\theta=\frac{3\frac{d}{\delta}-2\left(\frac{d}{\delta}\right)^2}{2\left[1-\left(1-\frac{d^2}{\delta^2}\right)^{3/2}\right]} \tag{32}$$

Eddy-Current Loss and Pf Formulas

Therefore the pf that the eddy-current losses reflect into the exciting winding is

$$\text{pf}=\cos\theta=\cos\tan^{-1}\frac{3\frac{d}{\delta}-2\frac{d^2}{\delta^2}}{\left[1-\left(1-\frac{d^2}{\delta^2}\right)^{3/2}\right]} \tag{33}$$

The surface resistance is the real part of the surface impedance, and is given by

$$r=\frac{16}{3\pi}\frac{1}{\gamma\delta}\left[1-\left(1-\frac{d^2}{\delta^2}\right)^{3/2}\right] \tag{34}$$

The loss per unit area N, given by equation 23 is thus

$$N=\frac{8}{3\pi}\frac{H_m{}^2}{\gamma\delta}\left[1-\left(1-\frac{d^2}{\delta^2}\right)^{3/2}\right] \tag{35}$$

Solid Iron Case

By solid iron it is meant here the case when the half thickness of the plate is equal to or greater than the depth of penetration. The moving surface as a function of time is given by equation 14

$$\zeta=\left(\frac{2H_m}{\omega\gamma B_0}\right)^{1/2}\sin\frac{\omega t}{2}\qquad 0<t<\frac{\pi}{\omega}$$

$$=\delta\sin\frac{\omega t}{2} \tag{36}$$

The electric field at the surface can be obtained from equation 20, realizing that now the fields penetrate up to the end of the half-period.

$$E=\frac{2H_m}{\gamma\delta}\cos\frac{\omega t}{2}\qquad 0<t<\frac{\pi}{\omega} \tag{37}$$

A plot of $\zeta(t)$, $E(t)$, and $H(t)$ is given in Fig. 7(B). Again the fundamental of E is obtained by Fourier series, whereby

$$E_1=\frac{8H_m}{3\pi\delta\gamma}(\cos\omega t+2\sin\omega t) \tag{38}$$

and the peak amplitude of E and H is given by

$$\bar{E}=\frac{8H_m}{3\pi\delta\gamma}(1-2j) \tag{39}$$

and

$$\bar{H}=-jH_m \tag{40}$$

This gives

$$N=\frac{8}{3\pi}\frac{H_m{}^2}{\gamma\delta} \tag{41}$$

and

$$\text{pf}=\cos\tan^{-1}\frac{1}{2}=0.894 \tag{42}$$

It can be seen from equations 41 and 42 that the power-loss per square meter, and pf for the solid case, can be readily obtained from equations 33 and 35 by substituting $d/\delta=1$. The loss per square meter is independent of the value of d, so long as d is greater than δ. Therefore, the maximum value of d/δ to be used in equations 33 and 35 is unity.

Search Coil Voltage

The variation of flux in the 1-inch iron-ring specimens (Fig. 8) on the basis of the theory presented here can be calculated as follows

Let n = number of laminations in the ring $=\frac{1}{2d}$

Considering the general case when $d<\delta$, let the separating surface be at a depth ζ from the outside surfaces of the laminations.

ϕ = the instantaneous value of flux in the iron ring

$$\phi=2nlB_0\left(d-2\delta\sin\frac{\omega t}{2}\right)\qquad 0<t<t_c \tag{43}$$

and

$$=2nlB_0d\qquad t_c<t<\frac{\pi}{\omega} \tag{44}$$

The voltage induced in a search coil of N turns wound around the iron ring is given by

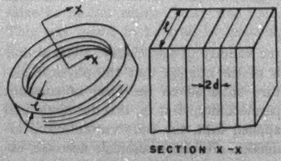


Fig. 8. Test specimen

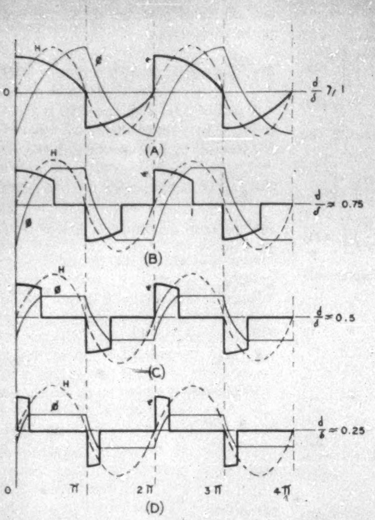


Fig. 9. Theoretical waveshape of e, ϕ, and H for various values of d/δ

$$e = -N\frac{d\phi}{dt}$$

$$= -2NnlB_v\delta\omega \cos\frac{\omega t}{2} \qquad 0<t<t_c$$

$$\text{or } \zeta<d$$

$$= 0 \qquad t_c<t<\pi/\omega \qquad (45)$$

The waveshapes of search coil voltage and induced flux therefore depend on the value of t_e and, hence, on the value of d/δ only. These waveshapes derived theoretically are shown in Fig. 9 for various values of d/δ.

The fraction of the half-period during which eddy currents flow is found to be:

$$= \frac{t_c}{T/2} = \frac{\frac{2}{\omega}\sin^{-1}\frac{d}{\delta}}{\pi/\omega} = \frac{2}{\pi}\sin^{-1}\left(\frac{d}{\delta}\right) \qquad (46)$$

This theory shows that in thin plates eddy currents do not flow continuously, but only for a fraction of the time during each half-cycle depending on the value of the fraction d/δ. This is contrary to the concept usually held, but an examination of the search coil voltages obtained in the tests will show that this is correct. The best proof of this theory lies in comparing the waveshapes of search coil voltages derived theoretically with those obtained experimentally.

Comparison of Theoretical and Experimental Waveshapes

Only a few of the more than 100 pictures taken experimentally are presented in Fig. 10, for comparison with those derived theoretically. Since the theoretical solution is based on an impressed mmf that is sinusoidal, every attempt was made to obtain sinusoidal mmf during the tests. To achieve this many practical problems were confronted, and that is why the experimental work was repeated three times before satisfactory conditions were obtained. In comparing the theoretical and experimental waveforms, it should be kept in mind that the actual magnetization curve is not exactly rectangular, and therefore no sharp corners should be expected in the experimental waveform of the search coil voltage.

An examination of Fig. 10 shows that the applied mmf was maintained essentially sinusoidal and that the theoretical and experimental waveshapes of search coil voltage correlate very well. Fig. 10(D) was purposely selected to show the maximum distortion that was tolerated in the applied mmf. Analysis showed that its third- and fifth-harmonic components were only 17.6% and 3%, respectively.

In Figs. 11(A) and 11(G), even though the peak applied mmf lies on the so-called linear portion of the magnetization curve, the waveform of search coil voltage is as predicted by the new theory, and not sinusoidal as given by the linear theory. The close correlation between the waveform of search coil voltages obtained experimentally and those predicted by this theory substantiates the validity of this theory. The author is of the opinion that the physical picture obtained on the basis of constant permeability, that is, that for sinusoidal applied field, the eddy currents and induction are also sinusoidal, decrease exponentially from the surface inward, and that eddy currents flow continuously, is

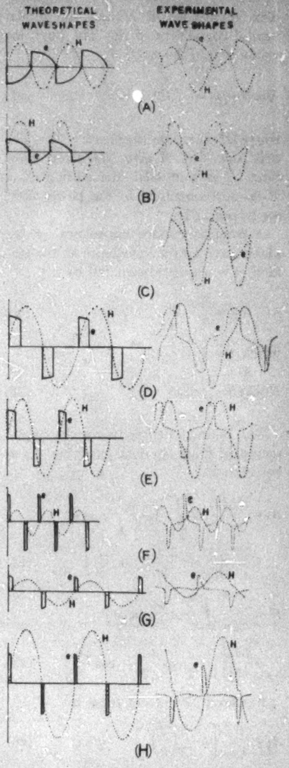


Fig. 10. Experimental and theoretical waveshapes of search-coil voltages

Theoretical e and H waves have been drawn, so that their amplitude and wavelength correspond to test values

	Ring, Inches	f	H_m	d/δ
A	solid	60	48	>1
B	1/4	120	229	>1
C	solid	120	145	>1
D	1/8	30	346	0.368
E	1/16	60	272	0.3
F	0.025	60	45	0.145
G	0.025	30	31	0.125
H	0.025	30	249	0.052

Vol. 7
Pt. 1
1959

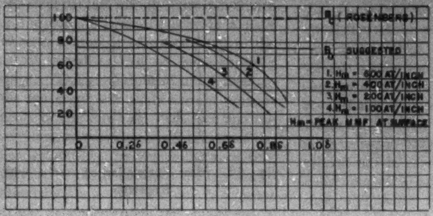


Fig. 11. Peak flux density at layers inside the material as a percentage of maximum density B_m at surface

far from correct and probably more misleading than helpful toward the actual solution of the problem.

Saturation Flux Density B_0

The value of B_0, the saturation flux density, to be used in any particular case is open to question. An assumption originated by Rosenberg[4] proposes that the value of B_0 should be that taken from the d-c magnetization curve of the material, corresponding to the peak field intensity H_m at the surface. Rosenberg made this assumption in order to make a simple calculation of eddy-current phenomena possible.

A careful examination of the problem will show why this assumption is not correct. If it is assumed that under dynamic conditions the induction density at any instant corresponds to the value of the magnetic-field intensity, given by the static magnetization curve of the material, then the value of B_0 will correspond to H_m at the outside surface layer only. Each succeeding inner layer is magnetized by a progressively smaller number of exciting ampere turns, because of the shielding effect of the eddy currents in the region between the outside surface and the layer under consideration.

Fig. 11 gives the peak flux density at various depths of the material and has been derived using Rosenberg's[4] eddy-current density curve and neglecting phase shift as suggested by him. Since, according to all eddy-current theories, the principal eddy-current losses occur in layers nearest the surface, where the phase shift is negligible, the shape of Fig. 11 would be essentially the same no matter which theory was used. This was found to be true by plotting Fig. 11 using other theories.

Fig. 11 shows that even for the case of strong magnetization (curve 1, for 600 ampere turns per inch) at depths 1/2 δ and 3/4 δ from the surface, the peak density reduces to about 80% and 60%, respectively, of the peak flux density at the surface. In the case of weaker magnetization, this reduction is even greater. An examination of Fig. 11 shows that some value of B_0 lower than B_m, the peak flux density at the surface, is more appropriate. It is recognized that an exact solution of the eddy-current phenomena is not possible, and that empirical calibration factors have to be determined to account for the departure from the ideal case assumed for analytical solution. It was found that the best value of B_0 to be used should equal 3/4 B_m, where B_m is the flux density corresponding to H_m, the peak value of mmf impressed at the surface, obtained from the d-c magnetization curve of the material. With the value of B_0, as Table I shows, the measured and calculated losses correlate very well over the entire range of tests performed. This value of B_0 is shown in Fig. 11 and clearly is more justified than B_m. The exact value of the fraction 3/4 can be open to question. The author believes that, since the shapes of B-H curves of most commonly used ferromagnetic materials do not differ materially, the fraction 3/4 will apply in almost all cases. The exact value of B_0 is, however, not too critical because the losses in solid iron are a function of $\sqrt{B_0}$.

For solid iron the power loss per square meter, assuming constant permeability, as derived by Ramo and Whinnery,[3] equation 1, page 241, is

$$N_1=\frac{1}{2}R_s|J_s|^2=\frac{1}{2}\left(\frac{\pi f\mu}{\gamma}\right)^{1/2}H_m^2$$

$$=\frac{1}{2}\left(\frac{\omega B_m/H_m}{2\gamma}\right)^{1/2}H_m^2 \text{ since } \mu=\frac{B_m}{H_m} \quad (47)$$

For a rectangular magnetization curve, the power loss, using $B_0=B_m$ as saturation flux density, is given by equation 41.

$$N=\frac{8}{3\pi}\left(\frac{\omega B_m/H_m}{2\gamma}\right)^{1/2}H_m^2=\frac{16}{3\pi}N_1=1.697N_1 \quad (48)$$

For a rectangular magnetization curve but using $B_0=3/4B_m$ as the value of saturation flux density, the power input formula becomes

$$N=\frac{8}{3\pi}\sqrt{\frac{3}{4}}\left(\frac{\omega B_m/H_m}{2\gamma}\right)^{1/2}H_m^2=1.47N_1 \quad (49)$$

Since the true magnetization curve of iron lies between the linear and rectangu-

Table I. Comparison of Computed and Experimental Values of Eddy-Current Losses

1-Inch Lamination			1/4-Inch Lamination			1/8-Inch Lamination			1/16-Inch Lamination		
H_{rms}	N_m	N_e	H_{rms}	N_m	N_e	H_{rms}	N_m	N_e	H_{rms}	N_m	N_e
30 Cycles											
27	184	190	32	285	252	32	227	235	31	100	74
66	790	790	80	1,167	1,045	64	605	480	62	162	130
173	3,470	3,500	162	3,410	2,990	150	1,036	915	125	213	217
247	6,200	6,160	221	4,700	4,070	245	1,250	1,210	243	280	307
60 Cycles											
34	410	390	32	386	350	33	325	365	32	260	206
75	1,360	1,350	74	1,415	1,310	74	1,280	1,310	65	463	370
174	5,040	5,030	174	5,430	5,000	131	2,670	2,150	136	685	613
247	8,700	8,730	238	8,350	8,250	251	3,980	3,450	248	840	886
120 Cycles											
22	250	285	36	620	590	32	440	500	27	362	365
69	1,730	1,675	80	2,460	2,115	74	1,730	1,830	88	1,500	1,254
102	3,250	3,120	162	7,090	6,300	184	7,820	7,700	194	2,360	2,125
137	5,070	4,900	240	12,900	11,800	251	11,400	12,400	248	2,540	2,500
240 Cycles											
36	812	860	34	835	800	39	875	950	37	788	890
79	3,075	2,900	79	3,215	2,920	72	2,385	2,520	75	2,620	2,510
176	11,150	10,240	142	8,430	7,300	135	7,000	6,770	99	3,930	3,400
247	19,480	17,450	234	18,500	16,000	235	16,940	15,730			
480 Cycles											
19	385	410	34	1,085	1,090	34	1,030	1,110	43	1,440	1,590
66	3,080	3,120	93	5,615	5,350	86	4,360	4,740	78	3,670	3,960
92	5,630	5,220	186	17,300	15,800	127	8,670	8,650	111	7,590	7,000
132	9,940	9,250	216	21,680	20,500	159	13,000	12,350	132	10,400	9,200
2,000 Cycles											
28	1,510	1,540	33	2,930	2,150	42	3,330	3,160	19	750	800
92	11,260	10,650	100	15,360	12,260	67	6,790	6,600	32	1,750	2,030
185	39,400	32,300	146	26,100	22,250	75	8,520	7,900	43	2,890	3,280

H_{rms} = impressed mmf in rms ampere-turns/inch
N_m = measured iron loss in watts/square meter
N_e = calculated value in watts/square meter

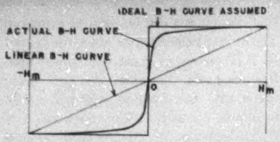


Fig. 12. True magnetization curve in relation to ideal and linear magnetization curves

lar case, as shown in Fig. 12, the rigorous solutions for the linear case as well as the rectangular case establish the limits within which the actual iron performs.

It is interesting to compare these results with that obtained by Rosenberg.[4] His formula is given in mixed units. When it is expressed in the units and notation of this paper, the following is obtained

$$N_{\text{Rosenberg}}=\left(\frac{4}{3}\right)\left(\frac{1}{2}\right)\left(\frac{\omega B_m/H_m}{2\gamma}\right)^{1/2}H_m^2$$
$$=1.333N_1 \quad (50)$$

Thus, Rosenberg's formula closely approaches equation 49 derived in this paper. This accounts for the fact that his formula has given good results, although it was based on intuition and not theoretically sound. His method does not give any information regarding the emf induced by the partial flux linkages, which are responsible for the reactance of the solid iron as seen from the exciting winding. The formulas derived in this paper are based on fundamental Maxwell's equations, and correctly explain all the phenomena associated with eddy currents. They have the further advantage of applying regardless of the thickness of the material.

Hysteresis

Hysteresis loss in solid iron has been a subject of much controversy. Some authors believe that in solid ferromagnetic materials the hysteresis loss is negligible compared with the eddy-current loss, and quite the contrary views are held by others.

When considering hysteresis, the ideal magnetization curve is that of a single crystal, shown in Fig. 13. The coercive force is shown as h_k. When hysteresis is neglected, the flux waves of saturation density B_0 penetrate the material from the surface when the field mmf H passes through zero. With hysteresis, the field mmf has to reach the value h_k, the coercive force, before the flux waves penetrate the material. The magnetic field h inside the material does not decrease from the value H at the surface to zero at the separating surface, see Fig. 13, but decreases to h_k. This results in an electric field and, hence, residual eddy currents of constant density in the region between the separating surface and d or δ, whichever is smaller. It has been shown by Haberland[1] that under these conditions the depth of penetration of flux is smaller because of the smaller effective field force, than in the case when hysteresis is neglected. Also, the peak value of the induced emf is smaller. According to Haberland,[1] the power loss due to eddy currents now becomes N_h = eddy-current loss per square meter taking hysteresis into account

$$=N\left(1-\frac{h_k}{4H_m}\right)$$

where N is power loss neglecting hysteresis.

The formula shows that, except for very low values of induction, the reduction in the eddy-current losses caused by hysteresis is small. This reduction is offset in some degree by the hysteresis loss.

The view that in massive iron hysteresis loss is negligible compared to eddy-current loss seems reasonable, because, except for very low values of induction, change of current amplitude is large for a small change of flux density amplitude, and eddy-current loss is a function of the former, and hysteresis loss of the latter.

Hysteresis loss was calculated for all the test points and found to be less than 4–5% except in the 1/16-inch laminations at low values of frequency and mmf. In view of the fact that an error of 10–15% is satisfactory in the estimation of iron losses, the effect of hysteresis in solid iron and thick laminations can be neglected. That is why in this paper test values are compared with calculated values which neglect hysteresis.

Effect of Temperature

The eddy-current loss formula, equation 35, for solid iron, that is, when $d/\delta \geqslant 1$ reduces to

$$N=\frac{8}{3\pi}H_m^{3/2}\left(\frac{\omega B_0}{2}\right)^{1/2}\sqrt{\rho} \qquad (51)$$

showing that in solid iron the eddy-current losses are proportional to the square root of the resistivity.

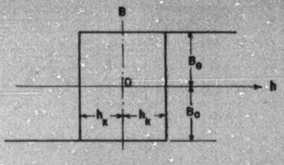


Fig. 13. Ideal magnetization curve with hysteresis

For very thin plates, that is, when $d/\delta << 1$, fourth and higher powers of d/δ can be neglected, giving

$$N\approx\frac{\sqrt{2}}{\pi}H_m^{1/2}(B_0\omega)^{3/2}d^2\frac{1}{\sqrt{\rho}} \qquad (52)$$

showing that for thin laminations the eddy-current losses are inversely proportional to the square root of the resistivity.

A series of tests were made on the ring specimens, which confirmed the temperature dependence of eddy-current losses as given above.

PF of the Iron Circuit

Although field theory is the foundation of all electromagnetic phenomena, it is very cumbersome in solving actual electromagnetic problems, except in the cases of simplest geometry. Most problems are conveniently solved by circuit theory and therefore, the proper values of resistance and reactance parameters are required to be used in the equivalent circuit, to take into account the effect of alternating field in iron. The proper value of R can be determined from the losses as calculated earlier, and an approximate value of power factor, and hence X, can be computed as shown in the following. Due to the nonlinearity of the magnetization curve of iron, both the voltage and the magnetizing current can not be simultaneously sinusoidal, and therefore it should be realized that in this case pf does not have its usual meaning.

Fig. 14 gives the approximate equivalent circuit and phasor diagram of the ring specimens on which iron loss tests were performed. The R_{h+e} branch is a resistive branch which is theoretically con-

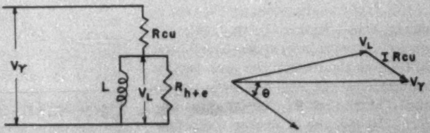


Fig. 14. Equivalent circuit and phasor diagram

nected in parallel with the pure inductance of the coil, in order to provide a path wherein a number of watts equal to the actual core loss can be hypothetically spent. The subscripts *h* and *e* refer to the hysteresis and eddy-current components of the core loss. The fact that the core loss in question is a function of the flux established by the inductance coil, together with the fact that the flux established by the inductance coil is to some extent dependent upon the magnitude of voltage drop across the ohmic resistance, dictates the circuit position of the hypothetical R_{h+e} branch.

$$V_L=\sqrt{(V_r-IR_{cu}\cos\theta)^2+(IR_{cu}\sin\theta)^2} \tag{53}$$

$$P_{\text{iron}}=W-P_{cu}$$

$$\text{Pf of the iron circuit}=\frac{P_{\text{iron}}}{V_L I}$$

The measured values of the power factor were determined by the foregoing equation. The analytical value of the pf is given by

$$\text{pf}=\cos\tan^{-1}\frac{3\dfrac{d}{\delta}-2\dfrac{d^3}{\delta^3}}{2\left[1-\left(1-\dfrac{d^2}{\delta^2}\right)^{3/2}\right]} \tag{33}$$

where

$$\delta=\left(\frac{2H_0\rho}{\omega B_0}\right)^{1/2}$$

The value of the pf depends on the value of d/δ only. When d/δ is equal to or greater than unity, the power factor is 0.894, which is amply supported by the test values as shown in Table II. As d/δ decreases, the pf also decreases according to the foregoing equation, and the measured and the test values are in reasonable agreement.

Table II. Comparison of Calculated and Experimental Values of PF

1-Inch Lamination			1/4-Inch Lamination			1/8-Inch Lamination			1/16-Inch Lamination		
H_{rms}	PF_c	PF_m	H_{rms}	PF_c	PF_m	H_{rms}	PF_c	PF_m	H_{rms}	PF_c	PF_m
30 Cycles											
27	0.894	0.85	32	0.894	0.9	32	0.86	0.86	31	0.462	0.52
66	0.894	0.825	80	0.894	0.825	64	0.68	0.75	62	0.35	0.397
178	0.894	0.795	162	0.845	0.81	150	0.47	0.5	125	0.26	0.243
247	0.894	0.795	221	0.75	0.765	245	0.37	0.354	243	0.19	0.17
60 Cycles											
34	0.894	0.93	32	0.894	0.91	33	0.894	0.82	32	0.64	0.693
75	0.894	0.8	74	0.894	0.82	74	0.855	0.815	65	0.482	0.476
174	0.894	0.79	174	0.894	0.855	131	0.694	0.705	136	0.35	0.30
247	0.894	0.86	238	0.894	0.805	251	0.52	0.481	248	0.26	0.2
120 Cycles											
22	0.894	0.81	36	0.894	0.815	32	0.894	0.815			
69	0.894	0.835	80	0.894	0.85	74	0.894	0.82	88	0.58	0.582
102	0.894	0.83	162	0.894	0.845	184	0.81	0.80	194	0.41	0.346
137	0.894	0.825	240	0.894	0.81	251	0.72	0.765	248	0.37	0.29
240 Cycles											
36	0.894	0.815	34	0.894	0.87	39	0.894	0.84	37	0.894	0.85
79	0.894	0.864	79	0.894	0.837	72	0.894	0.83	75	0.894	0.85
176	0.894	0.868	142	0.894	0.86	135	0.894	0.85	99	0.765	0.77
247	0.894	0.875	234	0.894	0.845	235	0.894	0.845			
480 Cycles											
19	0.894	0.87				34	0.894	0.88	43	0.894	0.87
66	0.894	0.81	93	0.894	0.845	86	0.894	0.805	78	0.894	0.84
92	0.894	0.855	186	0.894	0.895	127	0.894	0.847	111	0.894	0.888
132	0.894	0.88	216	0.894	0.83	159	0.894	0.825	132	0.894	0.87

H_{rms} = impressed mmf in rms ampere-turns/inch
PF_c = calculated value of pf
PF_m = measured value of pf

Table III. Iron Losses in 1/16-Inch Ring Specimen*

Frequency, Cps	Applied Mmf, H_{rms}	P, Dreyfus	P, Rosenberg	P, New Theory	P_{iron}, Measured
30	31.3	26	26.3	52.5	71
	61.7	31.8	31.8	92	116
	125.4	37.6	37.7	150	152
	184	40	40.6	184	178
	243	42.3	42.8	210	200
60	31.7	84	106	147	181
	64.5	119	129	263	330
	136	146	152	438	489
	190	162	163	540	536
	248	171	172	620	600
120	26.5	181.5	397	255	258
	55	370	507	630	761
	87.5	463	577	880	1,076
	132	555	615	1,110	1,346
	194	620	660	1,490	1,687
	248	655	690	1,730	1,810
240	37	515	†	643	571
	75	1,190	2,175	1,820	1,864
	99	2,170	2,320	2,550	2,771
480	43	880	†	1,160	1,055
	77.5	2,270	†	2,880	2,662
	111	3,670	†	5,300	5,462
	132	4,400	†	7,150	7,744
2,000	19	585	†	595	559
	32	1,000	†	1,510	1,310
	42.9	1,600	†	2,490	2,155

* The losses are calculated for the total volume of the ring, which is 578 cubic centimeters.
† Thickness is more than twice the depth of penetration and hence Rosenberg formula for losses in thin plates is not applicable.

Iron Losses in Thin Laminations

Table III presents a comparison of eddy-current losses in thin laminations calculated by the new theory with those given by the theories of Dreyfus[8] and Rosenberg.[4] These theories are discussed in brief in Appendix II. These losses, according to formulas given by Rosenberg and Dreyfus, are proportional to the square of the peak flux density. This would mean that as we increase the applied mmf, the peak flux density, and therefore the losses, will increase until the iron becomes saturated, and then any increase in applied mmf will produce little increase in losses. This is contrary to all test results since tests show that the losses are a function of applied mmf. It can be seen from Table III that Rosenberg and Dreyfus values have errors which might go as high as 400%. The calculated values by the new theory are within 15% of the test values, except for low frequency and low applied mmf when hysteresis loss, which has not been included here, is not negligible.

Experimental Work

The first core-loss tests on 1-inch, 1/4-inch, and 1/8-inch steel-bar specimens were made by the General Engineering Laboratory, of the General Electric Company, in Schenectady. Tests were made at 30, 60, and 120 cycles up to an applied mmf of about 400 ampere-turns per inch, using a 50-centimeter Epstein frame. In this setup, it was difficult to determine accurately the mmf acting on the specimens alone, because of the air gaps at the butt joints. It was therefore decided to prepare ring specimens.

The following rings of outside diameter

12 inches and inside diameter 10 inches were cut out of common steel plate, General Electric material *B4A26* in the "as received" condition:

1. One ring from 1-inch steel plate.
2. Four rings from 1/4-inch steel plate.
3. Eight rings from 1/8-inch steel plate.
4. Sixteen rings from 1/16-inch steel plate.

The 1/4-, 1/8-, and 1/16-inch rings were individually dipped in a fine solution of shellac in alcohol, thus providing a very thin coating of shellac as insulation between the laminations. Rings of same plate thickness were then assembled together to give four rings of cross-section 1 by 1 inch and outside diameter 12 inches and inside diameter 10 inches. Each ring was covered by electrical scotch tape no. *33* with half overlap. Two search coils of 30 turns each were then wound on each ring at opposite ends. Then each ring was covered with about 225 turns of uniformly wound no. 12 American Wire Gauge heavy, Formex-coated copper wire. This was covered with one layer of electrical scotch tape, and another coil of about 75 turns was wound uniformly around the whole ring. Splitting the magnetizing winding into two coils gave the flexibility of using 225±75 turns, to adjust to the voltage and current available.

A second series of tests were made at the Polytechnic Institute of Brooklyn to determine iron losses in the ring specimens at frequencies of 15, 30, 45, 60, 75, and 90 cycles. Because of inadequate capacity of the machine available, it was not possible to maintain sinusoidal current at high values of mmf.

A third series of tests were made on the ring specimens in the General Engineering Laboratory. These tests were made at 30, 60, 120, 240, 480, and 2,000 cycles. Since the smallest generator used was 60 kva, it was possible to use a larger noninductive resistance, which absorbed by far the major portion of applied voltage, and thus it was possible to force the current to be sinusoidal. Photographs of the wave forms of current and search coil voltage were taken simultaneously. Every care was taken to obtain all readings at about room temperature. D-c magnetic measurements were made on the rings by the Rowland ring method, using a ballistic galvanometer.

Conclusions

It has been shown that the eddy-current loss theory presented here, based on a rectangular magnetization curve and sinusoidal impressed mmf waveform, provides a simple loss formula that checks very well with test results over a very wide range of frequencies, flux densities, and lamination thickness. Although the actual magnetization curve of iron departs from the ideal shape assumed in the theory, comparison of analytical and experimental waveforms of search coil voltage indicates that magnetization of the iron does actually take place as predicted by the rectangular theory.

The oscillograms show that in thin laminations at high values of impressed mmf with sinusoidal wave form, the flux rises to its saturation value in the first few degrees of the time cycle, and, thereafter, the flux remains almost constant until the end of the half-cycle. Hence, the eddy currents that oppose the rise of flux are intermittent, being large in the first part of each half-cycle and nearly zero after the flux reaches saturation. Thus, the effect of applied mmf beyond that required to produce normal magnetic saturation is to increase the rate of change of flux, giving the same effect as an increase in frequency. This explains why the fundamental frequency core losses in rotating machines rise so rapidly (as the sixth to seventh power of flux, approximately), when the apparent flux density is increased beyond normal saturation limits.

The theory applies to rotating machines especially, because their mmf waves that create the fundamental flux have nearly sinusoidal waveforms. Since in transformers, and many other applications, the total flux, and not the applied mmf, is sinusoidal, a formula for eddy-current losses in this case using the idealized saturation curves has been derived in the thesis, although no comparison with test results was made.

As far as the author is aware, no satisfactory formula, analytical or empirical, for calculating eddy-current losses when the mmf wave is nonsinusoidal, is available today. It is intended, therefore, to extend this work to include the effects of harmonics in the applied mmf.

Appendix I

Losses by Time Average Dissipation

As indicated in the text, since H is sinusoidal, the harmonics of E do not contribute to the power flow into the metal. This, at first thought, seems to be wrong because for nonsinusoidal current flow the power loss is the sum of losses caused by the fundamental and the harmonics. Here the case is different. The wave form of E, the value of electric field e at the surface, is shown in Fig. 7(A). The value and waveform of electric field are different at different depths from the surface. The function $e(Z, t)$, can be easily obtained from Figs. 6 and 7(A), and is shown in Fig. 15. The electric field e has the value E after ζ exceeds Z; otherwise zero.

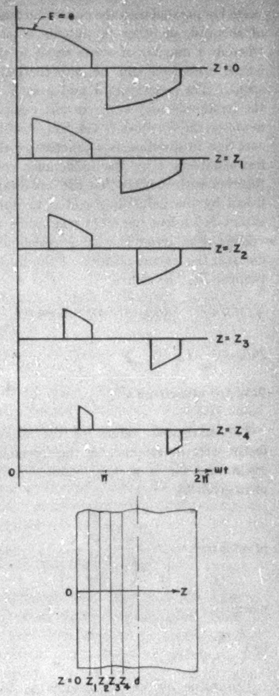


Fig. 15. Waveform of electric field at various depths from the surface

That the power loss is correctly given by equation 35

$$N=\frac{8}{3\pi}\frac{H_m{}^2}{\gamma\delta}\left[1-\left(1-\frac{d^2}{\delta^2}\right)^{3/2}\right]$$

can be verified by calculating the total dissipation

$$N=\int_0^d \gamma\langle e^2\rangle dz \tag{54}$$

where < > means time average.

$$\langle e^2\rangle=\frac{2}{T}\int_{\omega t=2\sin^{-1}z/\delta}^{\omega t=2\sin^{-1}d/\delta} E_{(t)}{}^2dt$$

$$\langle e^2\rangle=\left(\frac{2H_m}{\gamma\delta}\right)^2\frac{2}{T}$$

$$\int_{\omega t=2\sin^{-1}z/\delta}^{\omega t=2\sin^{-1}d/\delta}\cos^2\frac{\omega t}{2}\,dt$$

$$=\left(\frac{2H_m}{\gamma\delta}\right)^2\frac{1}{2\pi}\int_{2\sin^{-1}z/\delta}^{2\sin^{-1}d/\delta}\cos^2\frac{\theta}{2}\,d\theta$$

$$<e^2>=\left(\frac{2H_m}{\gamma\delta}\right)^2\frac{1}{2\pi}\left[2\sin^{-1}\frac{d}{\delta}+\right.$$

$$\left.\sin 2\sin^{-1}\frac{d}{\delta}-2\sin^{-1}\frac{z}{\delta}-\sin 2\sin^{-1}\frac{z}{\delta}\right]$$

$$=\left(\frac{2H_m}{\gamma\delta}\right)^2\frac{1}{2\pi}\left[2\sin^{-1}\frac{d}{\delta}+\right.$$

$$2\frac{d}{\delta}\sqrt{1-\frac{d^2}{\delta^2}}-2\sin^{-1}\frac{z}{\delta}-$$

$$\left.2\frac{z}{\delta}\sqrt{1-\frac{z^2}{\delta^2}}\right]$$

Substituting in equation 54

$$N=\gamma\left(\frac{2H_m}{\gamma\delta}\right)^2\frac{1}{2\pi}\int_0^d\left(2\sin^{-1}\frac{d}{\delta}+\right.$$

$$2\frac{d}{\delta}\sqrt{1-\frac{d^2}{\delta^2}}-2\sin^{-1}\frac{z}{\delta}-$$

$$\left.2\frac{z}{\delta}\sqrt{1-\frac{z^2}{\delta^2}}\right)dz$$

$$=\gamma\left(\frac{2H_m}{\gamma\delta}\right)^2\frac{1}{2\pi}\left[2d\sin^{-1}\frac{d}{\delta}+\right.$$

$$2d\frac{d}{\delta}\sqrt{1-\frac{d^2}{\delta^2}}-\left\{2z\sin^{-1}\frac{z}{\delta}+\right.$$

$$\left.\left.2\sqrt{\delta^2-z^2}-\frac{2\delta}{3}\left(1-\frac{z^2}{\delta^2}\right)^{3/2}\right\}_0^d\right]$$

$$=\gamma\left(\frac{2H_m}{\gamma\delta}\right)^2\frac{1}{2\pi}\left[\frac{4\delta}{3}+2d\frac{d}{\delta}\sqrt{1-\frac{d^2}{\delta^2}}-\right.$$

$$\left.2\delta\sqrt{1-\frac{d^2}{\delta^2}}+2\frac{\delta}{3}\left(1-\frac{d^2}{\delta^2}\right)^{3/2}\right] \quad (55)$$

$$N=\gamma\left(\frac{2H_m}{\gamma\delta}\right)^2\frac{1}{2\pi}\left[\frac{4\delta}{3}-2\delta\left(1-\frac{d^2}{\delta^2}\right)^{1/2}+\right.$$

$$\left.2\frac{\delta}{3}\left(1-\frac{d^2}{\delta^2}\right)^{3/2}\right]$$

$$=\gamma\left(\frac{2H_m}{\gamma\delta}\right)^2\frac{1}{2\pi}\left[\frac{4\delta}{3}-\frac{4\delta}{3}\left(1-\frac{d^2}{\delta^2}\right)^{3/2}\right]$$

$$=\frac{8}{3\pi}\frac{H_m^2}{\gamma\delta}\left[1-\left(1-\frac{d^2}{\delta^2}\right)^{3/2}\right] \quad (35)$$

Appendix II

The Theory of Dreyfus[6]

An important factor in the theory is the "reduced thickness" $\xi=\alpha h$ of the lamination, where h is the thickness of the lamination in millimeters.

$$\alpha=2\pi\sqrt{\nu\mu\lambda 10^{-5}} \quad (56)$$

ν=frequency of applied mmf
μ=permeability$=\frac{1}{0.4\pi}\frac{B}{\sqrt{2}H}$
B=flux density in gausses, corresponding to $\sqrt{2}H$
H=magnetic field intensity in ampere turns per cm (centimeter), rms
λ=conductivity in ohms per meter and millimeter

The distribution of eddy currents is similar for all laminations with the same reduced thickness ξ. This quantity has no dimension and is porportional to the ratio of the inductive to the ohmic resistance of the paths in which the eddy currents flow. Calculations show that ξ lies between 0.2 and π, and that the greater the value the more the field is concentrated toward the edges of the lamination, and the more the phase of the field in the middle lags behind the phase at the edge of the lamination. The eddy-current loss in watts per decimeter^3 is given as

$$\frac{W^{\mathrm{watt}}}{V^{\mathrm{cdm}}}=\left(\frac{\pi^2\lambda}{10\times 6}\right)f_\rho\left[\frac{\nu}{100}\frac{B_m}{1000}h^{\mathrm{mm}}\right]^2 \quad (57)$$

B_m is the mean flux density over the section of the lamination

$$B_m=B_{h/2}\frac{\sqrt{2}}{\xi}\sqrt{\frac{\cosh\xi-\cos\xi}{\cosh\xi+\cos\xi}} \quad (58)$$

where $B_{h/2}$=amplitude of flux density at the edges.

For small values of ξ

$$B_m\approx B_{h/2}\left[1-\frac{1}{3}\left(\frac{\xi}{2}\right)^4\right]$$

and for $\xi\geqslant\pi$

$$B_m\approx B_{h/2}\frac{\sqrt{2}}{\xi}$$

f_ρ is a factor to allow for the reaction of eddy currents on the inducing field, and is given by

$$f_\rho=\frac{3}{\xi}\frac{\sinh\xi-\sin\xi}{\cosh\xi-\cos\xi} \quad (59)$$

For $\xi<3.5$ we have approximately $f_\rho=1-0.0014\xi^4$ and for $\xi>3.5$ we have $f_\rho=3/\xi$. Curves giving the value of f_ρ and the ratio $B_m/B_{h/2}$ as a function of ξ are given for values of ξ up to 9.

The Theory of Rosenberg[4]

Rosenberg presented a separate formula for the eddy-current losses in "thin plates" whose thickness is less than twice the penetration depth, so that the total of their eddy currents is materially smaller than the exciting ampere-turns. The loss per cubic cm in watts is

$$L_{\mathrm{ccm}}=1.64(10^{-16})\frac{f^2B^2d^2}{\rho} \quad (60)$$

where

f=frequency of applied mmf
B=peak flux density at surface in lines per cm^2
d=lamination depth in cm
ρ=resistivity in ohms per cm and cm^2

References

1. Alternating Fields in Saturated Solid Iron (in German), G. Haberland, F. Haberland. *Archiv fur Elektrotechnik*, Berlin, Germany, vol. 19, 1930, pp. 126–33.

2. Theory of Strong Electromagnetic Waves in Massive Iron, W. MacLean. *Journal of Applied Physics*, New York, N. Y., vol. 25, Oct. 1954, pp. 1267–70.

3. Eddy-Current Phenomena in Ferromagnetic Materials, H. M. McConnell. *AIEE Transactions*, vol. 73, pt. I, July 1954, pp. 226–35.

4. Eddy Currents in Solid Iron (in German), E. Rosenberg. *Elektrotechnik Und Machinenbau*, Vienna, Austria, vol. 41, June 1923, pp. 317–25.

5. Fields and Waves in Modern Radio, S. Ramo, J. Whinnery. John Wiley & Sons, Inc., New York, N. Y., 1953.

6. Flux Distribution and Eddy-Current Losses in Armatures of Electric Machines (in German), L. Dreyfus. *Archiv fur Elektrotechnik*, vol. 4, 1915, pp. 99–139.

7. The Nature of Polyphase Induction Machines (book), P. L. Alger. John Wiley & Sons, Inc., New York, N. Y., 1951.

8. Short-Circuit Capabilities of Synchronous Machines for Unbalanced Faults, P. L. Alger, R. F. Franklin, C. E. Kilbourne, J. B. McClure. *AIEE Transactions*, vol. 72, pt. III, 1951, pp. 394–403.

9. Investigation of the Eddy-Current Anomaly in Electrical Sheet Steels, F. Brailsford. *Journal*, Institution of Electrical Engineers, London, England, vol. 95, pt. II, Feb. 1948, pp. 38–48.

10. Mathematical Description of Core Losses, J. W. Hale, F. R. Richardson. *AIEE Transactions*, vol. 72, pt. I, Sept. 1953, pp. 495–501.

11. Eddy-Current Losses in a Semiinfinite Solid Due to Nearby Alternating Current, H. Poritsky, R. P. Jerrard. *Ibid.*, vol. 73, pt. I, May 1954, pp. 97–106.

12. Solid Iron Conductors and Eddy Current Brakes (in German), E. Rosenberg. *Elektrotechnische Zeitschrift*, Vienna, Austria, vol. 44, 1923, pp. 1055–57.

13. Losses of Eddy Currents in Electric Brakes and Dynamos (in German), R. Rudenberg. Samm. Elek. Vortrage, Stuttgart, Germany, vol. 10, 1906, pp. 269–370.

14. Propagation Velocity of Ferromagnetic Domain Boundary, W. Shockley, C. Kittel. *Physical Review*, New York, N. Y., vol. 80, Dec. 1950, pp. 1090–94.

15. Losses in Electrical Sheet Steel, K. H. Stewart. *Journal*, Institution of Electrical Engineers, London, England, vol. 97, pt. II, 1950, pp. 121–25.

16. On the Heat Produced by Eddy Currents in an Iron Plate Exposed to an Alternating Magnetic Field, J. J. Thomson. *Ibid.*, vol. 28, 1892, p. 599.

17. Electromagnetic Fields, E. Weber. Vol. 1, John Wiley & Sons, Inc., New York, N. Y., 1950.

18. Field Transients in Magnetic Systems, E. Weber. *AIEE Transactions*, vol. 50, Dec. 1931, pp. 1234–46.

Discussion

Dr. Ernst Weber (Polytechnic Institute of Brooklyn, Brooklyn, N. Y.): Professor Agarwal's ingenious extension of Professor MacLean's earlier work on eddy-current losses in solid iron constitutes a very significant contribution in a field which was already intriguing to J. J. Thompson. The replacement of the magnetization curve by its rectangular idealization has been fully exploited here, and the reward is a rather clear and incisive picture of the peculiar distribution of eddy currents, limited in space in solid iron (MacLean), and limited in time in iron laminations where the depth of penetration is larger than half the thickness (Agarwal). In combination, the result is surprising agreement between measured and calculated values of eddy-current losses over wide ranges of frequency and magnetization values. Of course, one must accept a heuristic equivalent field value 0.75 $B_{\max}$ which cannot be theoretically deduced but which apparently serves to account for the finite slope of the magnetization curve above the knee of saturation.

It is surprising that agreement between measured and calculated losses exists down to H_{max} of the order of 20 ampere-turns per inch; however, almost all types of steel sheets and plates have their knees of saturation below this value of magnetization except cast iron. It would be interesting to check the theory in case of cast iron and also down to 5 and 10 ampere-turns per inch, though this would be more of academic interest than of practical value.

Another point of interest would be the case of pulsed magnetic excitation such as often found in magnetic amplifiers where the knowledge of losses is of considerable importance.

P. L. Alger (Rensselaer Polytechnic Institute, Troy, N. Y.): This paper marks a very important advance in our knowledge of eddy-current behavior in laminated steel. The paper shows that the eddy currents that occur in the laminations of rotating machines are intermittent, instead of being continuous as always assumed heretofore. That is, the eddy currents which oppose the flux change in a particular lamination rise rapidly at the start of each half-cycle, then decrease to a very small value during most of the cycle, and rise again at the end of the half cycle.

This occurs because the mmf impressed on the magnetic circuit of a rotating machine varies sinusoidally in time, in accordance with the sinusoidal distribution of the current around the periphery of the machine. Thus, for a given high value of flux density, the impressed mmf at the start of a cycle is much larger than necessary to produce a sinusoidal flux. These conditions are very different from those in a transformer, where the exciting current has a highly peaked waveform, and the mmf is, therefore, low at the start of each cycle.

The formulas developed in accordance with this new understanding are extremely simple and practical, while at the same time they cover the full range of frequency, steel thickness, and flux density, so long as the mmf has a sinusoidal waveform in time.

The discusser is sure that the ideas presented in the paper, when combined with the physicist's ideas of domain theory, will lead to further important advances in the understanding of eddy-current phenomena in transformers and all sorts of apparatus.

G. Angst (General Electric Company, Schenectady, N. Y.): This paper represents a valuable contribution to the theory of eddy-current calculation in solid iron and in laminated magnetic structures. The worth of the paper is enhanced by the close agreement between calculations and test results.

It has been known to designers of rotating electric machinery for some time that the classical formula for eddy currents in laminated cores breaks down when the magnetic flux density is increased beyond normal peak values of about 100 kilolines per square inch. Experimental evidence has shown that the fundamental frequency eddy-current losses increase with the sixth power of the flux density instead of the second power. The author's analysis predicts and explains this phenomenon satisfactorily. It is believed that the theory, when applied to high-frequency core losses due to harmonics, should also result in closer agreement between prediction and test.

In the field of solid-iron application the author's study has contributed considerably to a better understanding of the physical phenomena involved. Doctor Agarwal has limited himself to the solution of the two-dimensional problem. In rotating machines with solid iron rotors, end effects may become important. This then presents a three-dimensional field problem and it is hoped that further studies will be made in this area, leading to a satisfactory solution.

E. F. Hansen (General Electric Company, Schenectady, N. Y.): The author has made an outstanding contribution to electrical engineering by developing and presenting this new and more accurate method of predicting eddy-current core loss in ferromagnetic materials as used in rotating electric machinery.

Any engineer concerned with design and performance testing of electric machinery will testify that the need for better methods of calculating core loss goes back many decades. The laboratory engineer or scientist active in development or testing of magnetic materials for rotating machinery will also be aware of the lack of correlation between core loss measured in the laboratory and core loss measured on actual tests of machines. The new method, properly applied, should help to remove those errors in the calculation of core loss attributable to eddy-current loss at fundamental frequency of mmf.

This paper will very likely be of practical use not only in calculating core losses of rotating electric machinery but also in other laminated magnetic circuits such as solenoids and a-c lifting magnets. Other indirect applications would be where eddy currents in solid or laminated iron dictate the time constants of field windings. There are other cases where eddy currents in rotating solid iron pole faces serve the useful function of damping out torsional vibrations. The author mentioned several other applications.

A quite different application would be the induction heating of ferromagnetic materials. The sine wave of applied mmf would be representative of most core-type induction heating fixtures.

The inherent accuracy of the new method is demonstrated by excellent agreement with the author's experimental test results. This is even more gratifying when compared to the substantial errors possible when calculations were made using the previously available methods.

It seems to the discusser that the author's work also demonstrates the way in which progress often results from a change in point of view. Many engineers have always thought of eddy-current loss as a function of flux density when in reality there seems to be considerable advantage in considering it as a function of magnetizing current. This concept is multiplied in importance because the sine wave of mmf is more likely to occur in magnetic circuits with air gaps than the sine wave of flux. Since the air gap is a significant factor in rotating electric machinery, the concept of eddy-current loss as a function of a sine wave of mmf is quite appropriate.

On the other hand, since pure sine waves of mmf would be rare, the author's proposal to investigate the effect of harmonic currents is to be encouraged. A further suggestion, for the benefit of those of us who for so long have been thinking in terms of sine waves of flux, would be to further clarify the relationship of the losses encountered with the sine wave of current as compared to the sine wave of flux. This paper should be an inspiration to encourage students to enter the field of electrical power engineering. There are still many, many more new areas of discovery not yet uncovered.

H. Poritsky (General Electric Company, Schenectady, N. Y.): The type of analysis given in this paper should be welcomed by the electrical industry. The analysis is based on the arbitrary assumption that the flux density is constant and equal to $3/4B_m$. Yet the fair agreement of the predicted losses with measured values serves to justify the assumptions made.

Generally, the flux which penetrates a magnetic lamination or a solid portion of a motor or generator is not driven by a uniform, tangential magnetic field. The flux may take a path which is partly in the metal, partly in air, and the metallic portion may be subject to different degrees of saturation. An analogous situation is given by a ring in which the windings are not uniformly distributed, and the ring is not complete.

Even if one simplifies the problem to that of a uniformly continuous winding over a solid ring, the winding to be fed directly by a generator, the applied mmf will be sinusoidal only if there is a large external resistance in the circuit. If the resistance of the winding and generator impedance are small, then the emf will be sinusoidal. In intermediate cases the voltage equation for the circuit of the winding may result in a more complicated condition at the ring surface than either sinusoidal mmf or sinusoidal emf.

For the foregoing reasons, it is desirable that a direct solution of Maxwell's equations with general boundary conditions be worked out. If E denotes the induced electric field, then Maxwell's equations lead to simultaneous equations (in centimeter-gram-second practical units)

$$\frac{\partial B}{\partial t}=10^{-8}\frac{\partial E,}{\partial z}, H=H(B) \tag{61}$$

$$\frac{\partial H}{\partial z}=\frac{4\pi}{10}I=\frac{4\pi}{10}\lambda E \tag{62}$$

or eliminating E

$$\frac{\partial B}{\partial t}=4\pi\lambda 10^{-9}\frac{\partial^2 H}{\partial z^2} \tag{63}$$

For a sinusoidal mmf the value H at $z=0$ is sinusoidal in time. For sinusoidal emf E at $z=0$ is sinusoidal in time. If there is a significant resistance in the winding circuit, then the boundary condition involves both H and E at $z=0$.

The solution of differential equations 61 and 62 is quite involved analytically. This discusser has tried several methods of approach, all of which turned out to be complex. He has also used direct numerical methods based upon introducing a finite grid of points in the (z, t)-plane, at the same time replacing the derivatives by finite differences. He hopes to present these methods at a future occasion.

Both the analytical and the numerical methods are time consuming. Yet with the advent of high-speed computing machines they become economically feasible.

Frederick J. Young (Carnegie Institute of Technology, Pittsburgh, Pa.): The author uses Maxwell's equations in conjunction with the idealized magnetization curve of Fig. 1 to derive expressions for eddy current losses in iron. This assumption, if taken literally, leads to an inconsistency. The tangential component of E at the moving surface is discontinuous. This consequence of the idealized treatment has been recognized by McConnell.[1]

Really McConnell's earlier treatment (vaguely mentioned in the author's reference 3) is remarkably similar to that of this paper in the matter of assumptions, methods, and conclusions. Perhaps the contribution of the author should have been stated more explicitly.

REFERENCE

1. EDDY-CURRENT PHENOMENA IN FERROMAGNETIC MATERIALS, H. M. McConnell. *AIEE Transactions*, vol. 73, pt. I, July 1954, pp. 226–34.

Paul D. Agarwal: The interest shown in this paper by the discussers is very encouraging and the author wishes to thank them for their contribution. It is gratifying to note that they recognize the value of the formulas and the data presented in this paper, and the need for further work in the field.

Dr. Weber's suggestion of checking the theory for values of magnetization below the knee of the saturation curve presents a formidable difficulty. At these low values of magnetization the hysteresis loss is predominant, and the division of the total measured losses into eddy-current losses and hysteresis losses depends on the empirical formulas used. The accuracy of these formulas, and hence the division of experimental losses, is questionable. Therefore, comparison of eddy current-losses at these low densities loses significance. Dr. Weber's idea of extending this work to magnetic amplifiers is an excellent one.

Mr. Alger's suggestion of combining this work with the physicist's ideas of domain wall theory of magnetization is indeed valuable. The author has already been working on this possibility and hopes to present some results in the not-too-distant future.

Eddy-current losses in laminated cores, according to most formulas in use today, are proportional to the square of the peak flux density. This would mean that as the applied mmf is increased, the peak flux density and hence the eddy current losses, will increase until iron is saturated. Beyond this any increase in applied mmf will produce little increase in losses. This is contrary to all test results presented in the paper, which show that losses are a function of applied mmf. Mr. Angst points out that the theory presented in this paper satisfactorily explains the fundamental frequency eddy-current losses in rotating machines, which experiments show, rise as the sixth to seventh power of flux density when flux density is increased beyond 90–100 kilolines per square inch. Solid-iron rotors in which end effects are not minimized by either copper end rings or extension of rotors beyond the stator stacks present a difficult problem. The extension of the theory to this three-dimensional field problem would indeed be a useful investigation. At this time the author wishes to acknowledge the benefit that he has derived from his discussions with Mr. Angst during their association at the General Electric Company.

Mr. Hansen points out how progress often results from a change in point of view. The theory presented here has brought out two new concepts which are contradictory to views presently held.

1. That eddy-current losses should be considered a function of mmf and not of flux density.

2. That in thin sheets eddy currents flow intermittently and not continuously.

Both the new concepts are, however, correct, as amply supported by experimental data. The author is glad Mr. Hansen mentions the need for encouraging students to enter the field of electrical power engineering. From the author's point of view this cannot be stressed enough. The number of young men entering this field has been decreasing rapidly with every year and the situation has become alarming.

Dr. Poritsky's discussion gives an opportunity to elaborate on some of the points brought out in the paper. It is recognized that as we penetrate into the material, each succeeding inner layer is magnetized by a progressively smaller number of exciting ampere turns because of the shielding effect of the eddy currents in the region between the outside surface and the layer under consideration. But because of the assumed ideal magnetization curve the flux density wave has the same magnitude irrespective of the magnitude of mmf. It is the departure of the actual B-H curve from the ideal assumed that has necessitated the factor 3/4 as explained in the section "Flux Density B_0" and in Fig. 10. That the magnetization does seem to take place as if waves of constant flux density were penetrating the material is substantiated by comparison of experimental and analytical waveshapes of search coil voltages presented in Fig. 11. The author agrees with Dr. Poritsky that in many applications neither the applied mmf nor the total flux is sinusoidal and it is pointed out in the conclusions of the paper that the work needs to be extended to include the effects of harmonics in the applied mmf. A rigorous solution of Maxwell's equations for nonsinusoidal mmf and using the B-H relation presented by the actual hysteresis loop will be a great analytical accomplishment. However, the practical value of such a solution is questionable. Since in samples from the same material the value of H for the same value of B may vary $\pm 15\%$, it is the author's opinion that an exact solution of the problem with its enormous complexity is not warranted.

Table IV. D-C Magnetization Data on 1/4-Inch Ring Specimen

Resistivity, 14.4 Ohm-Cm

H, Ampere-Turns/Inch	B, Kilolines/Inch^2
8	44.0
16	71.5
28	84.5
40	91.6
80	100.0
120	104.6
160	107.5
240	112.7
320	116.6
400	119.2

That the assumption of the idealized magnetization curve leads to discontinuous E at the separating surface, though obvious from the field pictures presented, needs further explanation according to Mr. Young's discussion. The separating surface is at the wavefront $h=0$, at which point the permeability is infinite according to Fig. 1. This is a technique often followed in an effort to simplify the solution; i.e., the problem is first idealized, and then the solution is examined for corrections required because of departure from idealization. That is why it is mentioned in the paper that "In comparing the theoretical and experimental waveforms, it should be kept in mind that the actual magnetization curve is not exactly rectangular, and therefore no sharp corners should be expected in the experimental waveforms of the search coil voltages." This immediately removes the inconsistency mentioned by Mr. Young.

Mr. Young's attention is directed to the last paragraph in the opening paragraphs, where it is clearly stated that several authors (including McConnell) have solved the solid-iron case under the assumption of a rectangular magnetization curve. It was not considered necessary to name each author. The last sentence in the opening section reads "However, so far as the author is aware, no previous writer has derived the formulas for the laminated iron case that are developed here. Nor have any authors obtained satisfactory checks between calculated and test results over so wide a range of applied mmf, frequency, and plate thickness as are covered here." Mr. Young should have found this an explicit statement of the author's contribution.

During the presentation of the paper at the Winter General Meeting of the AIEE in New York, information was requested on the magnetization curve of the ring specimens used. Although the B-H curves of various specimens were slightly different, the values given in Table IV are fairly representative.

A Method for Calibration of Precision Voltage Dividers

CHARLES B. PINCKNEY
ASSOCIATE MEMBER AIEE

ONE PHASE of electrical measurements which is rapidly becoming more important and requiring higher degrees of accuracy is the measurement of voltage ratio. Conventional 4-arm bridge techniques are often used where two impedances are connected in series, energized from either or both of the voltages whose ratio is being measured, and adjusted so that the voltages across them exactly match the two being compared, both in magnitude and phase angle. The voltage ratio can easily be computed if the characteristics of the impedances are known. Sze has developed a system which uses this technique and will measure voltage ratio with an accuracy of about 50 parts per million (ppm) and phase angle to 0.2 minute at frequencies up to 10,000 cycles per second (cps).[1]

A device which lends itself well to measurements of ratio is the voltage divider, often referred to as a potentiometer. According to strict definition, it is not a potentiometer, but usage of the term is fairly common and is used here for clarity. In any case, it is a tapped impedance, presenting a voltage at its output terminals which is equal to or less than that applied to its input terminals, depending upon the position of the tap. This position may be calibrated to indicate the ratio of the output to the input voltage; a decimal fraction or percentage scale is often convenient. Such a scale must be evaluated and any resulting corrections taken into consideration if the voltage divider is to be used for accurate measurements. It should be noted that varying degrees of loading at the output terminals will affect these corrections, and a common practice is to specify open-circuit conditions.

If the voltage divider has continuous resolution or is adjustable in very small finite steps, it is obviously impracticable to provide corrections for every possible reading of the scale. It is often satisfactory to calibrate only the points which are used most frequently, or to calibrate the main, or cardinal points, 0.1, 0.2, 0.3, etc. The corrections may be determined in two ways: set the divider on the exact point to be calibrated and adjust the voltage ratio to coincide, or measure a known fixed voltage ratio with the divider. In either case, the correction to be applied to the divider is the difference between the true ratio and that indicated by the divider.

The purpose of this paper is to describe a technique which may be used to establish known fixed ratios on a voltage divider, which may then be used as a standard for the calibration of others. The method was investigated using a type of voltage divider which is basically an autotransformer, but the theory applies equally well to resistive and capacitive dividers.

Theoretical Considerations

A perfect voltage divider providing ratios in steps of tenths would consist of 10 sections with identical impedances. The absolute values of these impedances need not be known, but each section would have a relative value of exactly 0.1 If the divider was not perfect, the relative values would differ more or less from 0.1, but in any case their sum must equal unity, since the output cannot exceed the input voltage. The ratio presented at any of the taps is the sum of the relative values of the sections involved.

The impedance ratio which can be measured most accurately is that of unity, that is, when the two impedances are identical. This principle may be used to determine the relative values of the voltage divider sections, using a variation of a technique which was developed at least as long ago as 1915.[2] A precision ratio transformer was developed by B. Diener, of Hughes Aircraft Company, also using a variation of this technique, in an independent study. If the ratio of each section to the section immediately below it in the divider is measured, section 10/section 9, section 9/section 8, etc., the results may be combined by a simple algebraic manipulation to express each section in terms of a common reference, say section 1. For example

$$\frac{\text{section 4}}{\text{section 1}}=\left(\frac{\text{section 4}}{\text{section 3}}\right)\left(\frac{\text{section 3}}{\text{section 2}}\right)\left(\frac{\text{section 2}}{\text{section 1}}\right)$$
$$=K_4$$
$$\text{section 4}=K_4(\text{section 1})$$

As previously stated, the sum of the relative values of the sections must equal unity.

$$\text{section 1}+\text{section 2}+\ldots+\text{section 10}=1$$

Substituting for each section its equivalent value in terms of section 1

$$\text{section 1}+K_2(\text{section 1})+\ldots+K_{10}(\text{section 1})=1$$

Solving for the relative value of section 1

$$\text{section 1}=\frac{1}{1+K_2+\ldots+K_{10}}$$

The relative value of each section may be calculated from the relative value of section 1 and the corresponding value of K.

If the standard divider is constructed so that section 1 is further subdivided into 10 smaller sections, these sections can be evaluated in the same manner and used to provide voltage ratios from 0.01 through 0.09. In this case, however, the sum of the relative values is equal to the relative value of section 1 rather than unity. Theoretically, this process could be carried on indefinitely, but in practice the reduction in allowable voltage (and the corresponding decrease in sensitivity) limit the calibrations to the second or third decades for most cases.

The validity of the measured relative values of the sections can be tested by using various combinations of sections to produce a ratio approaching unity. The measured ratio of (section 1+section 2)/(section 3+section 4) should equal the ratio of this combination as calculated from the relative values. This applies to any number of sections when compared to an equal number of adjoining sections.

Other ratios may be set up for measurement by using less than 10 sections for the standard divider, recalculating the relative values accordingly. If 9 sections are used, ratios in multiples of 1/9 are established, 0.1111111, 0.2222222,

---

Paper 59-22, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted February 10, 1958; made available for printing November 14, 1958.

CHARLES B. PINCKNEY is with the Hughes Aircraft Company, Culver City, Calif.

The author wishes to express his appreciation to the Primary Standards Laboratory, Hughes Aircraft Company, for the facilities made available and for encouragement in the preparation of this paper.

Vol.
Pt. 1
1959

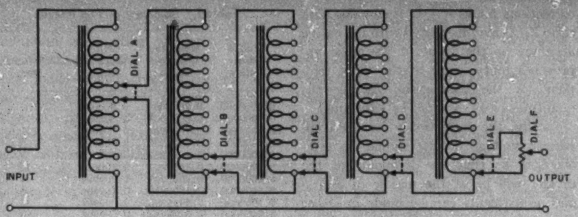


Fig. 1. Diagram of ratio transformer type of voltage divider

0.3333333, etc. This procedure also serves as a cross check for the relative values, since certain ratios can be set up by using different combinations of sections. A ratio of $^1/_3$ can be established by using 3 sections of a 9-section divider, 2 sections of a 6-section divider, or 1 section of a 3-section divider.

The possibility exists that the ratios presented by the standard will be too high or too low to be used for calibrating the desired point on another divider. For example, the sum of the relative values of 6 of the standard sections might be 0.5999980, which could be too low to be measured using the 0.6 setting of the divider to be calibrated. If necessary, the characteristics of the standard can be altered by adding a small series impedance to either end and including it in the relative value measurements. If impedance is added to the bottom of the standard divider, it will be included in the relative value of section 1 and all the taps will be shifted upward in value.

The foregoing arguments are based entirely on considerations of impedance magnitude. While this analysis is valid for measurements made with direct voltage, the quantities involved with alternating voltage are related in phase as well as magnitude and must be treated accordingly. Excessive phase shifts in voltage dividers are usually undesirable, however, and a well-designed divider will have little or none within the frequency range of normal use. Typical values for the units used in this investigation are of the order of 0.1 minute at 1,000 cps. The error in magnitude introduced by adding phasors directly, without regard to phase angle, is less than one ppm for angles smaller than 3 minutes. A phase angle will cause a reduction in bridge resolution, however, since the quadrature component of the unbalance signal is still present in the detector after the real component is balanced, and this may affect the amount of phase shift which can be ignored.

Description of Tests

Voltage Dividers Used

Voltage dividers of the type known as ratio transformers were used in all of the tests performed. They are basically autotransformers, with a switching arrangement based on a variation of the Kelvin-Varley circuit, which provides continuous adjustment through a succession of paralleling decades; see Fig. 1. The Kelvin-Varley circuit applies to resistive voltage dividers and requires 11 sections per decade, with the paralleling decades shunting 2 sections at a time so as to form 10 equivalent sections. The ratio transformer uses 10 sections per decade. The windings are on a common core, and the paralleling decades contain the same number of turns as the section across which they are connected, so that a voltage is induced in the windings of the paralleling decades equal to that in the bridged section and no loading takes place. A separate core is used for the fourth and fifth decades to reduce the number of turns necessary for the first decade, but the loading effect of the second core seems to be negligible.

The validity of the arguments which have been presented was tested by calibrating a number of voltage dividers by intercomparison of their component sections. In the discussion which follows, when the term "standard" voltage divider is used, the reference is to these units. The standard dividers were used to determine corrections for a similar divider which had been calibrated by the National Bureau of Standards using Sze's apparatus and technique; this unit is referred to as the "test" voltage divider. The divider used for calibration of the standards is designated the "comparison" divider; see Fig. 2.

Calibration of Standard Dividers

The ratio of each section in the standard to the section immediately below it was measured with the comparison divider, as shown in Fig. 2, and these measurements used to calculate the relative value of each section in the manner which has been described. The final accuracy of the calibration of the standard is largely dependent upon the accuracy with which the ratios of adjacent sections can be measured. The following discussion indicates that the closer the standard is to being perfect (all sections identical), the more accurate the calibration of it.

The circuit diagram of the ratio transformers used in these tests is shown in Fig. 1. The ratio of output to input voltage is read directly from the dials of the switches controlling the various paralleling decades. Dials A, B, C, D, and E are scaled from 0 to 9. Dial F is scaled from 0–100. If the corrections for the dials are designated a, b, c, d, e, and f, the true ratio for any setting is given by

$$R_t = 0.1(A+a)+0.01(B+b)+0.001(C+c)+0.0001(D+d)+0.00001(E+e)+0.0000001(F+f) \quad (1)$$

If the ratio transformer is used to measure the ratio of Z_1 to Z_2, as shown in Fig. 3, and Z_1 and Z_2 are identical, the resulting bridge circuit will balance at the same setting of the ratio trans-

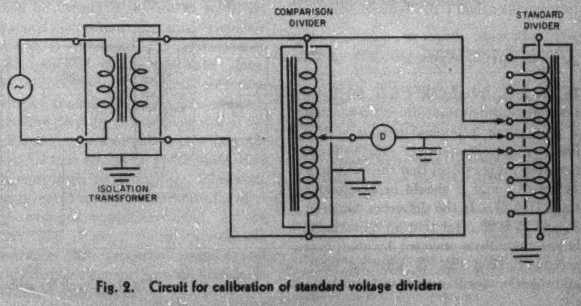


Fig. 2. Circuit for calibration of standard voltage dividers

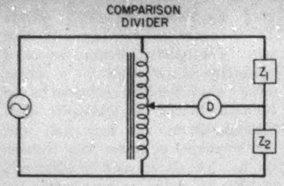


Fig. 3. Measurement of the ratio Z_1/Z_2

former whether Z_1 and Z_2 are connected as shown or interchanged, since the setting represents the exact midpoint of the comparison divider. If Z_1 and Z_2 have slightly different values, the divider will require adjustment for balance on direct and reversed connections. Assuming for the moment that the adjustment is small enough to be made entirely with dial F, the settings and corrections of dials A, B, C, D, and E will remain the same and may be combined for convenience.

Let

R = combined settings of dials A, B, C, D, and E
r = combined corrections a, b, c, d, and e
F, F' = settings of dial F for direct and reversed connections
f, f' = corrections to dial F at F and F'

With Z_1 and Z_2 connected as shown in Fig. 3

$$\frac{Z_1}{Z_2}=\frac{1-(R+r+F+f)}{(R+r+F+f)} \qquad (2)$$

With Z_1 and Z_2 interchanged:

$$\frac{Z_1}{Z_2}=\frac{(R+r+F'+f')}{1-(R+r+F'+f')} \qquad (3)$$

Typical values for the quantities in equations 2 and 3 are:

$R=0.50000$
$r=3(10^{-6})$
$F,F'=5(10^{-6})$
$f,f'=1(10^{-7})$

If equations 2 and 3 are averaged and terms smaller than $1(10^{-10})$ discarded:

$$\frac{Z_1}{Z_2}=\frac{R}{R+(F+f)-(F'+f')} \qquad (4)$$

In the ratio transformers used for these tests, the adjustment controlled by dial F is a 5-ohm 3-terminal resistor with a linearity specification of 1%. From equation 1 it is evident that the maximum values of f and f' should be no greater than $1(10^{-7})$. If the difference between Z_1 and Z_2 is so great that an adjustment larger than that provided by dial F is necessary, then dial E must be used, and equation 4 will contain the additional corrections e and e'. It is obvious that ratio Z_1/Z_2 may measured with increasing accuracy as the value of the ratio approaches unity, since at the exact midpoint of the comparison divider the quantities $(F+f)$ and $(F'+f')$ are equal.

Ideally, the correction r for the comparison voltage divider should be −0.0000050, so that the midpoint of the divider will be at the middle of dial F and adjustments can be made in either direction for direct and reversed connections of Z_1 and Z_2. The reading corresponding to the midpoint will be 0.5000050, and the ratio Z_1/Z_2 can deviate from unity by as much as 20 ppm and still be measured using only dial F. The midpoint of the comparison divider used in these tests was at a setting of about 0.5000033, which provided sufficient leeway. If necessary, this midpoint can be shifted by the addition of impedance to either end of the comparison divider.

The standard dividers were removed from their cases so that connections could be made to the various sections; these taps are available at the switch contacts. Contact resistance at these points will affect the measurements and care must be taken to avoid this source of error. It was also found that the calibration of a standard is affected slightly by the position of the paralleling sections, possibly due to distributed capacitance effects. Thus, the corrections obtained for a standard divider are invalid at the conventional output terminals, since the only way to obtain the desired ratio at these terminals is to move the paralleling sections. The calibration is valid at the switch contacts as long as the paralleling sections are in the position of calibration, wherever it may have been. Data taken for one of the standard dividers indicates that the calibration of a given point was affected by 20 ppm or more by the position of the paralleling decades.

Table I. Summary of Test Results

Voltage Divider Setting	Measured Voltage Ratio	NBS Certified Ratio	Difference, Ppm
0.9000000	0.9000004	0.90000	<11
0.8000000	0.8000006	0.80000	<13
0.7000000	0.7000003	0.70000	<15
0.6000000	0.5999990	0.600000	2
0.5000000	0.4999967	0.500000	7
0.4000000	0.3999951	0.400000	12
0.3000000	0.2999961	0.299994	7
0.2000000	0.1999972	0.199996	6
0.1000000	0.0999984	0.099997	14
0.0900000	0.0899984	0.089997	16
0.0800000	0.0799982	0.079998	3
0.0700000	0.0699979	0.069997	13
0.0600000	0.0599978	0.059997	13
0.0500000	0.0499979	0.0499980	2
0.0400000	0.0399984	0.0399988	10
0.0300000	0.0299987	0.0299991	13
0.0200000	0.0199993	0.0199996	15
0.0100000	0.0099992	0.0099995	30

POWER SOURCE

Power for the measurements was supplied from a stable audio oscillator through a shielded and balanced isolation transformer. All of the measurements listed in Table I were made at 1,000 cps; this was the specified upper frequency limit of the test voltage divider. Exploratory data were taken at frequencies of 60 and 400 cps, and the results compared favorably with those observed at 1,000 cps. Voltages of about 2 volts per standard divider section were used, but it was found that voltage level had little effect on the results as long as the manufacturer's ratings were not exceeded.

DETECTOR SYSTEM

The detector system consisted of a battery-powered amplifier with a voltage gain of more than 10,000, the output of which was applied to an oscilloscope for visual presentation. Extreme selectivity was obtained in the amplifier by the use of twin-T filters. Phase shift in the voltage dividers reduce the resolution to about 2 ppm in extreme cases, but this was compensated to a large extent by using a technique of adjusting the voltage divider for equal amounts of

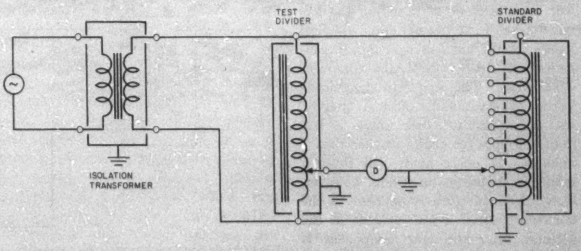


Fig. 4. Circuit for calibration of test voltage divider

unbalance signal on either side of the minimum and taking the midpoint of these settings as the null value. Tests were made using a mutual inductor to introduce a quadrature voltage into the detector circuit so that a complete balance could be reached, but the results were substantially the same, and it was felt that the system should be kept as simple as possible with a minimum of equipment. Over-all resolution of better than 0.5 ppm was achieved.

SHIELDING AND GROUNDING

Coaxial leads were used for all connections in the circuits. Various systems of shielding were tried, as was a Wagner ground. The Wagner ground had no apparent effect on the results and was not used for most of the tests. In the final tests, shielding was confined simply to grounding the cases of all equipment. Where the voltage dividers used as standards were removed from their cases to make connections to the taps, the shields around the cores and windings were grounded.

Results of Tests

In order to test the repeatability of the system, six different units were calibrated at 1,000 cps for use as standard voltage dividers. These were used to determine corrections for a test voltage divider which had been calibrated at the National Bureau of Standards (NBS) using Sze's apparatus and technique. The results for the six separate tests were in agreement with each other within eight ppm or better, and the deviation of their average from the calibration performed at the NBS is shown in Table I. The over-all accuracy of 0.005% for Sze's circuit is a general specification covering voltage ratios from 1 to 0.01 and frequencies up to 10,000 cps, and it is thought to be much better under some of the conditions for which these data are compared. For example, at 1,000 cps, the accuracies for ratios of 0.1 and 0.5 are believed to be 20 and 5 ppm, respectively.

Investigations are in progress at the NBS which will enable voltage ratios to be measured with much greater accuracies than are now possible. A capacitive voltage divider will be used in conjunction with a ratio transformer especially constructed by the NBS and accuracies of better than one ppm are anticipated.

It is hoped that when these facilities are available, they will provide a more precise evaluation of the system just described.

Conclusions

A system has been developed for the calibration of precision voltage dividers, based on the extreme accuracy with which equal or nearly equal impedances may be compared. A voltage divider is calibrated by intercomparison of its component sections and then used as a standard for the calibration of other dividers. Investigations were made at frequencies up to 1,000 cps, and accuracies of 0.01% or better were obtained for voltage ratios from 0.01 through 0.09 in steps of hundredths and from 0.1 through 0.9 in steps of tenths. See Fig. 4.

Calibrations performed on a voltage divider agreed with data furnished by the NBS within 30 ppm or better. The accuracy of the NBS calibration is estimated to range from 5–50 ppm, depending on the exact ratio.

References

1. MEASUREMENT OF VOLTAGE RATIOS AT AUDIO FREQUENCIES, Wilbur C. Sze. *AIEE Transactions*, vol. 76, pt. I, Sept. 1959, pp. 444–49.

2. THE TESTING OF POTENTIOMETERS, F. Wenner, E. Weibel. *Bulletin*, National Bureau of Standards, Washington, D. C., vol. 11, 1915, p. 15.

Evaluation of Solderless Wrapped Connections for Central Office Use

S. J. ELLIOTT
ASSOCIATE MEMBER AIEE

SINCE 1952, when the solderless wrapped connection first was described publicly,[1] its use in telephone central offices has grown steadily. Several hundred million solderless connections are being wrapped annually in the Bell System today, and the growth is continuing.

The tangible and immediate results of solderless wrapping have been gratifying. For example, the use of solderless wrapping has reduced manufacturing costs by speeding up many wiring operations and by reducing troubles caused by wire clippings and solder splashes. Furthermore, since solderless wrapping avoids the risk of damaging heat-sensitive insulation by soldering operations, it has made widespread use of plastic insulated wire practicable, and this is leading to substantial additional savings.

In the end, however, these savings could be illusory if the use of solderless wrapped connections degraded telephone service or increased the maintenance effort required in the telephone plant. The laboratory evalution of solderless wrapped connections has been continued, therefore, in order to assess the risk of deterioration in service and to provide guidance for the design of connections that are most likely to be reliable. This work has revealed certain limitations of solderless wrapped connections, but, at the same time, it has provided considerable assurance that properly designed and properly made connections will be reliable in central office service.

Many persons have inquired about the methods used for evaluating the capabilities of solderless wrapped connections and about the results that have been obtained since publication of earlier papers.[2,3] Since the inquiries continue undiminished year after year, it appears that there is sufficient interest in solderless wrapping to warrant another paper on the subject. An attempt is made in this paper therefore to bring the story on evaluation up to date.

Description of Connection

A solderless wrapped connection is made by wrapping a wire tightly around a

---

Paper 59-33, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 29, 1958; made available for printing November 20, 1958.

S. J. ELLIOTT is with the Bell Telephone Laboratories, Inc., New York, N. Y.

Many persons, working in several closely knit teams, have participated in the development of solderless wrapped connections. Each of those persons has played an important part in the development. In particular, the author wishes to acknowledge the contributions of H. L. Coyne and R. H. Van Horn, who supervised the several phases of the evaluation program.

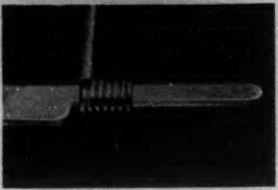


Fig. 1. Typical solderless wrapped connection

terminal, and the connection is held together thereafter by the elastic stresses left in the two members. A typical solderless wrapped connection is shown in Fig. 1.

For Bell System applications, a minimum of 5 turns of wire is specified when no. 22 gage wire is used, and a minimum of 6 turns is specified when no. 24 gage wire is used. The wire should be closely wound on the terminal, but turns of wire should not overlap.

Only solid copper wire has been approved for solderless wrapping. The use of stranded wire presents a number of difficulties and has not been investigated in detail.

RELIABLIITY OBJECTIVES

The general reliability objective in the development of solderless wrapped connections has been that the connections should remain mechanically secure and electrically stable during manufacture, shipment, and installation and for 40 years thereafter in telephone central office service.

The mechanical security objective cannot be defined in absolute terms, for too little is known about the magnitudes and distributions of the disturbances to which wires are subjected in service. As a comparative objective, however, solderless wrapping should not increase the occurrence of broken wires and loose connections beyond the levels that now prevail with soldered connections.

The electrical stability objective can be stated more specifically. Not more than one connection in 10,000 should exhibit resistance fluctuations greater than 0.1 ohm in service. The electric noise produced by resistance fluctuations of this magnitude is considered to be at the threshold of transmission impairment in the most critical transmission circuits now in service.

Electrical Measurements

The method used for checking the electrical stability of solderless wrapped connections has been described fully by Van Horn.[2] In essence, it consists of a resistance measurement made by the familiar voltmeter-ammeter method. The voltage drop across the connection is measured with a moving coil millivoltmeter while a direct current of 0.1 ampere flows through the connection.

The measurement of interest is the variation of resistance when the connection is disturbed mechanically. The disturbance is created by plucking the terminal, that is, by slowly deflecting the free end of the terminal a predetermined distance and then releasing it abruptly, allowing the terminal and the connection to vibrate freely. The resistance variation (ΔR) is the difference between the maximum and minimum resistance value observed during the disturbance.

Although the moving coil millivoltmeter is too sluggish to follow rapid fluctuations of resistance, the measurement is surprisingly sensitive. In a series of measurements made by an appropriate electric noise meter, while connections were subjected to vibration of the sort encountered in service, the measured noise levels consistently were lower than the levels calculated from the results of the simple ΔR measurements. It was concluded that the simple ΔR measurement would be adequate for routine development tests, and, since it could be made far more rapidly than any of the more refined measurements that had been explored, it was adopted.

ANALYSIS OF ΔR MEASUREMENTS

Before reviewing the accelerated aging tests that have been employed in the evaluation of solderless wrapped connections, it is necessary to describe the method that will be used to summarize the results. This method, a product of hindsight rather than foresight, has been used for only a short time in the solderless wrapped connection development.

A major problem in the evaluation, of course, has been to demonstrate by means of comparatively small samples that the probability of ΔR exceeding 0.1 ohm is no more than one in 10,000. The expense of testing large enough samples to determine the frequency distribution of ΔR in every case would have been prohibitive. Over a period of years, however, a moderately large number of tests were made on a few particular types of connections. The cumulative sample sizes in those cases, although still small compared with 10,000, were large enough to warrant attempts at curve fitting.

As measured by the chi-square test, the expression that seems to fit the observed distributions best is obtained by first grouping the ΔR data in cells as follows:

Cell Number x	ΔR, Ohms	Number of Connections in Cell
0	$0 \leq \Delta R \leq 0.001$	n_0
1	$0.001 < \Delta R \leq 0.01$	n_1
2	$0.01 < \Delta R \leq 0.1$	n_2
3	$0.1 < \Delta R \leq 1.0$	n_3
⋮	⋮	⋮

The arithmetic mean of the grouped distribution then is calculated as follows:

$$m = \frac{0n_0 + 1n_1 + 2n_2 + 3n_3 + \dots}{n_0 + n_1 + n_2 + n_3 + \dots}$$

$$= \frac{n_1 + 2n_2 + 3n_3 + \dots}{N} \tag{1}$$

where N is the total number of connections in the sample.

Once the mean m has been calculated, the probability of finding a connection in cell x can be expressed as

$$P_x = \frac{m^x \epsilon^{-m}}{x!} \tag{2}$$

Bearing in mind that $0! = 1$, the probabilities associated with the first few cells become

$$P_0 = \epsilon^{-m} \tag{3A}$$

$$P_1 = mP_0 \tag{3B}$$

$$P_2 = mP_1/2 \tag{3C}$$

$$P_3 = mP_2/3 \tag{3D}$$

The reader may recognize that equation 2, the general expression for P_x, is the same as the expression for the Poisson distribution. The physical interpretation, however, is different. In the Poisson distribution, P_x is the probability that x defectives will occur in a random sample of size N if m is the expected average number of defectives in a sample of that size. As a description of the ΔR distribution, however, P_x is the probability that ΔR, for a single connection chosen at random, will fall between the limits defined by the cell number x. For a random sample of N connections, then, the number of connections expected to fall in cell number x will be NP_x.

In general, the agreement between equation 2 and observed ΔR distributions has been best for the more stable types of connections, types in which high values of ΔR rarely occur and for which m is small. These, of course, are the types of connections that are desirable. For less stable types of connections, the types that would be rejected as unreliable, the agreement has been poorer. The dividing line between good agreement and poor agree-

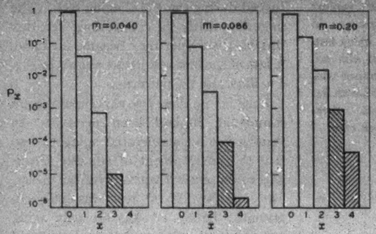


Fig. 2 (left). Examples of ΔR distribution defined by equation 2. Shading indicates cells for which ΔR exceeds 0.1 ohm

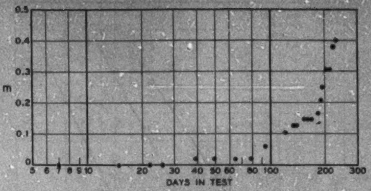


Fig. 3. Effect of 105 C accelerated aging test on solderless connections wrapped with no. 24 gage copper wire on solder-dipped twin-wire terminals of wire spring relay. Sample size 48

ment appears to be somewhere in the vicinity of $m=0.3$.

The mean m is a convenient statistic to use in summarizing the results of a group of ΔR measurements, and it will be used for that purpose in the discussion of accelerated aging tests. Qualitatively, low values of m indicate stable connections, and high values indicate unstable connections. Quantitatively, in those cases where m is less than about 0.3, m defines the observed ΔR distribution very effectively.

The distribution corresponding to $m=0.086$ (shown in Fig. 2) is of particular interest. It represents the case for which there is one chance in 10,000 that ΔR will exceed 0.1 ohm. Consequently, a value of $m=0.086$ in service is the maximum value that will satisfy the stability objective for central office use of solderless wrapped connections in the Bell System.

One more point needs to be made before proceeding to the discussion of accelerated aging tests. The variance of the distribution described by equation 2 is equal to the mean m. It follows, then, that the standard deviation is equal to the square root of m. This relationship will be used later in developing criteria for deciding whether or not connections are stable enough to be approved.

Accelerated Aging Tests

Fairly early in the development of solderless wrapped connections, it was concluded that electrical instability, if it occurred at all, would result principally from relaxation of stresses in the wire and terminal, the stresses that hold the connection together. Since that time, all aging tests employed in the evaluation of solderless wrapped connections have included elevated temperatures to accelerate the relaxation process.

The work of Mason and others[2,3] indicated that the stress in copper wire (measured with respect to its value one day after a connection was wrapped) would relax about 40% in 40 years at room temperature. To allow for some error in the room temperature prediction, and to allow for the fact that temperatures in enclosed equipments may approach 55 C (degrees centigrade) in some cases, it has been assumed that the stresses in solderless connections wrapped with copper wire may relax as much as 50% under central office conditions. That has been the degree of stress relaxation aimed at in the various accelerated aging tests.

One of the early tests consisted simply of baking the connections for 3 hours at 175 C. (Mason had shown that the stresses in connections wrapped with copper wire would relax 50% in about $2^1/_2$ hours at 175 C, and the extra half-hour was needed to bring the specimens from room temperature up to the oven temperature.) At the end of the heat treatment, the connections were cooled to room temperature and then were checked for electrical instability. The general experience was that any solderless wrapped connection which was stable before the heat treatment still would be stable after the heat treatment.

Although such results were encouraging, they were, at the same time, disconcerting. For example, some of the connections that behaved so well in the 175 C test were wrapped on terminals that scarcely could be considered good mechanical structures for supporting stresses over long periods. The twin-wire terminal of the wire-spring relay, in particular, fell in the questionable category. At that time it consisted simply of two parallel nickel-silver wires which were bonded together by dipping them in soft solder and which then were serrated. The parallel wires by themselves did not have sufficient torsional stiffness to support the stresses that are required to hold a solderless wrapped connection together. And soft solder, because of its cold flow properties, is a notoriously poor material for supporting stresses. It was doubtful that the soft solder, at room temperature, could maintain the stresses long enough at levels high enough for solid-state diffusion to occur. Since 175 C was not far below the softening temperature of the solder, there was at least a suspicion that something akin to a soldered joint had been produced in the accelerated aging test.

There were questions, also, about the metallurgical behavior of the copper wire at the elevated temperature. It was known, for example, that recrystallization is more likely to occur in copper at 175 C than at lower temperatures.

In short, it was suspected that a temperature as high as 175 C might do more than accelerate the aging process: it might alter the physical nature of the aging process itself. If this were so, then the 175 C test might give a false picture of the aging that would occur in actual service.

105 C AGING TEST

It was decided therefore that some exploratory aging tests should be run at a substantially lower temperature. For several reasons, a temperature of 105 C was chosen. It was well below the softening temperature of tin-lead solders, it was low enough so that there should be little likelihood of recrystallization occurring in the copper wire, yet it was high enough to produce the desired stress relaxation in a few months. Mason's work had indicated that the stresses in the wrapped connections would relax 50% in about 150 days at 105 C, so the test period was set at 150 days, or about 5 months.

As compared with the desired service life of 40 years, the 3-hour 175 C test represented a 100,000 to 1 acceleration of the stress relaxation process. Since the 5-month 105 C test would represent an acceleration of only 100 to 1, it was expected to be a far more realistic aging test.

Five months is a long time, however, to wait for test results. It was decided

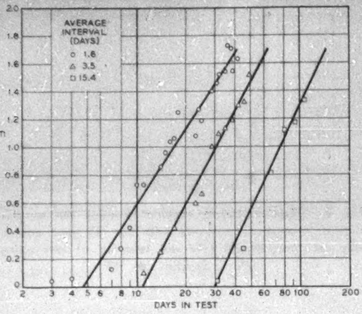


Fig. 4. Effects of varying frequency of disturbance in 105 C accelerated aging test. Connections wrapped with no. 24 gage tinned copper wire on 0.010- by 0.062-inch flat nickel-silver terminals. Sample size 48 in each case

therefore that the test connections should be measured periodically during the aging test so that any instability would be detected as soon as it occurred. It is important to note that this decision introduced two more changes in the accelerated aging test: 1. it subjected the connections to temperature cycling, for they were removed from the oven and allowed to cool whenever they were measured, and 2. it subjected the connections to mechanical disturbances while they were being aged, for the terminals were plucked whenever resistance variations were measured.

The first connections subjected to the 105 C aging test were wrapped on the solder-dipped twin-wire terminals that were described earlier. Connections of that type had survived the 175 C test without showing any evidence of instability. In the 105 C test, however, they soon began to exhibit measurable resistance fluctuations, and they grew more and more unstable as the test continued. The history of that first group of connections is plotted in terms of the statistic *m* in Fig. 3. It was evident that the 105 C aging test with periodic cycling and measurement was more severe than the undisturbed 175 C test.

Subsequently an experiment was performed to compare the relative effects of temperature, temperature cycling, and mechanical disturbance. One group of connections was aged at 105 C, was cooled to room temperature once each week, and was measured (plucked) while at room temperature on alternate weeks. At the end of 150 days, the value of *m* for that group was 2.1.

A second group of similar connections was aged at 105 C, was cooled to room temperature once each week, but was not disturbed by measurements. After 150 days, *m* was 0.2.

A third group was aged continuously at 105 C without either temperature cycling or mechanical disturbance. After 150 days, *m* was 0.04.

A fourth group was maintained continuously at room temperature but was measured every two weeks. After 150 days, *m* was 0.02.

Three important inferences were drawn from the results of this experiment:

1. Although solid-state diffusion may occur during undisturbed aging of solderless wrapped connections, thus tending to offset the detrimental effects of stress relaxation, repeated mechanical disturbances during the aging process can impede diffusion and can lead to unstable connections.

2. Since connections may be disturbed from time to time in service, accelerated aging tests should include periodic mechanical disturbances.

3. Temperature cycling alone can produce a form of mechanical disturbance if the temperature coefficients of expansion of wire and terminal are not equal.

Another related experiment was performed to study the effects of the frequency with which connections are disturbed during the 105 C aging test. The more frequently the connections were disturbed, the sooner they became unstable. A few of the test results are shown in Fig. 4 to illustrate the pattern.

Long before these two experiments were completed, it had been necessary to standardize an accelerated aging test so that the development of suitable terminals could proceed without further delay. The 150-day 105 C test was adopted, primarily because it was the only aging test up to that time that had revealed highly significant differences among various types of connections, differences that usually were consistent with qualitative analyses of the various mechanical structures. The connections were cooled to room temperature once each week, and their resistance variations were measured at room temperature every other week. Each connection was plucked two or three times during the ΔR measurement. Eventually a plucking amplitude of 1/32 inch was standardized.

A standard procedure for the preparation of test specimens also evolved gradually. It has become the usual practice now, in preparing each group of test connections, to use two wrapping bits (one that wraps more tightly than the average bit, and one that wraps less tightly), two wrapping tools (one power-driven and the other manually operated), two grades of wire (one comparatively hard and the other comparatively soft), and two operators. All 16 combinations of the 4 parameters are included at least twice in each test nowadays.

RESULTS OF 105 C ACCELERATED AGING TEST

Altogether, more than 11,000 solderless wrapped connections have been subjected

Table I. Results of Accelerated Aging Tests on Solderless Connections Wrapped with Tinned Copper Wire

Type of Terminal	Material	Finish	105 C Aging Test				175 C Screening Test			
			No. 22 Wire		No. 24 Wire		No. 22 Wire		No. 24 Wire	
			N	m	N	m	N	m	N	m
Rectangular:										
0.045 by 0.045	Brass	Electrotin	136	0	296	0	168	0	328	0.009
0.0319 by 0.062	Brass	Electrotin	152	0.026	152	0.020	160	0	160	0
0.0319 by 0.062	Nickel silver	None	152	0	104	0	976	0.15	2050	0.078
0.0253 by 0.062	Nickel silver	None			24	0.053	105	0.37	93	0.13
0.0201 by 0.062	Nickel silver	None			144	0.084	116	0.88	342	0.16
0.0159 by 0.062	Nickel silver	None			24	1.2			121	0.38
0.0126 by 0.062	Nickel silver	None			24	1.8			203	0.40
0.010 by 0.062	Nickel silver	None			96	2.1			1164	0.42
Embossed:										
0.0253 by 0.062	Nickel silver	None					136	0.31	137	0.007
0.0201 by 0.062	Nickel silver	None			64	0			274	0.16
0.0159 by 0.062	Nickel silver	None							131	0.084
0.0126 by 0.062	Nickel silver	None			64	0			131	0.015
0.010 by 0.062	Nickel silver	None			64	0			180	0.11
Wire Spring Relay:										
Single wire	Nickel silver	None	144	0.12	288	0			144	0.010
Twin wire	Nickel silver	None	144	0.021	513	0.006			749	0.064

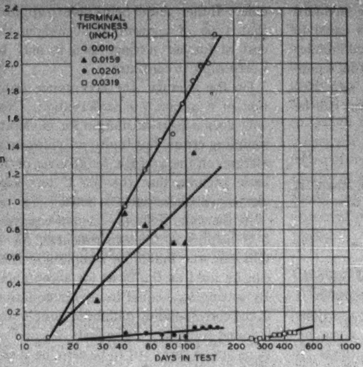
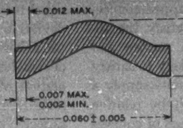


Fig. 5 (left). Results of 105 C accelerated aging tests. All terminals flat nickel silver, 0.062 inch wide. Connections wrapped with no. 24 gage tinned copper wire



NOMINAL THICKNESS OF STOCK (INCH)	DIMENSION A (INCH)
0.010	0.021 ± 0.002
0.013	0.023 ± 0.002
0.016	0.025 ± 0.002
0.020	0.026 ± 0.002

Fig. 6 (right). Embossing approved for Bell System terminals

to the 105 C accelerated aging test. Most of the tests have been aimed at the problem of finding terminal configurations which would yield reliable connections and which, at the same time, could be manufactured economically. The results of a number of those tests are summarized in Table I along with the results of a screening test that is described in a later section.

Fig. 5 shows the results that were obtained with no. 24 gage tinned copper wire wrapped on terminals punched from several widely used thicknesses of nickel-silver stock. The performance of the thinner terminals was considered to be unsatisfactory. Various types of longitudinal embossing were investigated to find means for improving the thin terminals, and the form shown in Fig. 6 finally was standardized. Although this form is not necessarily optimum, it is a convenient form to manufacture; and it has behaved well in the accelerated aging test, as shown in Table I.

The early single-wire terminal of the wire-spring relay was formed from a round wire by flattening, serrating on one side, and solder dipping. Its performance was unsatisfactory, but it was improved simply by omitting the solder. Its present form is shown in Fig. 7.

The twin-wire terminal of the wire-spring relay was more of a problem. Many designs were conceived and investigated, but most of the designs that showed promise from the solderless wrapping standpoint were objectionable from the manufacturing standpoint. In the end, a compromise design was adopted. The twin wires are twisted tightly together, then they are coined in a closed die which has a trapezoidal cross section. The resulting terminal is shown in Fig. 8.

Although most of the 105 C tests have been made with no. 24 gage wire, a number of tests have been made also with no. 22 gage wire. As shown in Table I, the results indicate that no. 22 gage connection the heavier terminals are as stable as no. 24 gage connections but that no. 22 gage connections on the lighter terminals are inferior.

The aging tests of no. 26 gage connections are not completed yet. The preliminary results indicate, however, that no. 26 gage connections, even when wrapped with as many as 9 turns of wire, are less stable than 6-turn no. 24 gage connections on the same types of terminals.

For certain types of wiring in the Bell System, it has been standard practice to use untinned copper wire. Solderless connections wrapped with untinned wire, however, tend to deteriorate sooner than connections wrapped with tinned wire. Fig. 9 illustrates results obtained with untinned wire.

The data that have been presented so far should be sufficient to provide a bird's-eye view of the results that have been obtained in the 105 C aging test. Other alleys and byways have been explored, but the picture would not be sharpened perceptibly by presenting more details. It is more pertinent at this point to consider what use can be made of the test results.

Criteron for Acceptance (105 C Aging Test)

The purpose of the accelerated aging test, of course, is to distinguish between

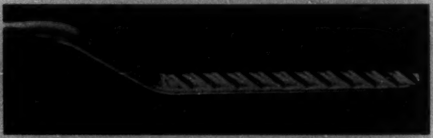


Fig. 7. Single-wire terminal of wire-spring relay

Fig. 8. Twin-wire terminal of wire-spring relay

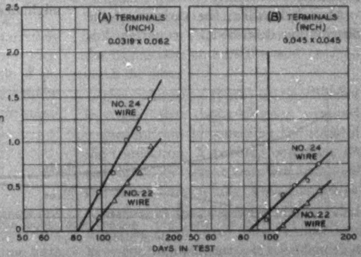


Fig. 9. Effect of 105 C accelerated aging test on connections wrapped with untinned copper wire on electrotinned brass terminals. Sample size 32 in each case

those types of solderless wrapped connections which are likely to satisfy the 40-year stability objective in service and those types which are not likely to satisfy the objective. The test results supply a reasonably clear picture of the relative instability of the several types of connections, but this by itself is not enough. Somewhere on the instability scale, the development engineer eventually must draw a line and say, at least to himself, "I will approve the use of types of connections that fall below this line; I will not approve types that fall above it." In other words, he must establish a criterion for acceptance.

The criterion for acceptance has not remained static as the development of solderless wrapped connections progressed. Instead, it has been revised several times as the aging test evolved and as the information obtained from aging tests expanded. Its present form has considerably more meaning and is more convenient to use than some of the earlier forms.

The criterion for acceptance is based upon two premises: 1. that the 105 C accelerated aging test does, in fact, produce at least as much instability as 40 years of central office service will produce, and 2. that equation 2 is an adequate description of the ΔR distribution. Although final confirmation will not be available until about 1990, evidence that these premises are valid is increasing year after year.

It was stated previously that the value of m should not exceed 0.086 in service if the resistance variation of not more than one connection in 10,000 is to exceed 0.1 ohm. If the 105 C test is an adequate simulation of service conditions, then those types of connections which have values of m below 0.086 in the 105 C test should be acceptable. This would be the criterion for acceptance if very large samples were tested. Where small samples are tested, however, it is prudent to make allowance for sampling errors.

As stated earlier, the variance of the distribution described by equation 2 is equal to m. Consequently, the standard deviation of the distribution is equal to the square root of m. If a large number of samples of size N are drawn from a population whose true mean is m', then roughly 5% of the sample means can be expected to fall below the value

$$m_{0.05} = m' - 1.645\sqrt{m'/N} \qquad (4)$$

A reasonable acceptance level, then, is

$$\begin{aligned} m_{0.05} &= 0.086 - 1.645\sqrt{0.086/N} \\ &= 0.086 - 0.482/\sqrt{N} \qquad (5) \end{aligned}$$

In other words, the stability of the connections will be considered acceptable if the mean m of a sample of N connections is less than the value of $m_{0.05}$ given by equation 5. The acceptance level is plotted as a function of N in Fig. 10.

As an aid to decision making, it also is possible to set up a criterion for rejection. If a large number of samples of size N are drawn from a population whose true mean is m', then roughly 95% of the sample means can be expected to fall below the value

$$m_{0.95} = m' + 1.645\sqrt{m'/N} \qquad (6)$$

A reasonable rejection level for the case $m' = 0.086$, then, is

$$m_{0.95} = 0.086 + 0.482/\sqrt{N} \qquad (7)$$

In other words, the stability of the connections will be considered unacceptable if the mean m of a sample of N connections exceeds the value of $m_{0.95}$ given by equation 7. Further testing of that particular type of connection is not very likely to be profitable, so it might as well be rejected.

Establishing a rejection limit that does not coincide with the acceptance limit provides a zone of uncertainty in which the connections are neither clearly acceptable nor clearly unacceptable. It recognizes the possibility that further testing of a marginal type of connection might demonstrate that the type is acceptable. If the value of being able to approve that type of connection outweighs the cost of further testing, then it may be worth while to continue.

On the average, the acceptance criterion of equation 5 is conservative. Not only does it provide margin for moderate sampling errors, but it also provides some margin, on the average, for an error in the basic premise that the 105 C test adequately simulates aging in service.

At the same time, the form of equation 5 is helpful to the experimenter, for it tells him the minimum sample size that can be used as the basis for acceptance. By setting $m_{0.05}$ equal to zero in equation 5, the corresponding minimum sample size is found to be 32. If he tests 32 connections and m turns out to be zero, he is entitled to approve the connections without further testing. With a smaller sample, he would not be entitled to approve them even though m turned out to be zero.

From a practical standpoint, it is prudent to test a larger sample when approval is needed in the shortest possible time. If ΔR for even a single connection exceeds 0.001 ohm in a sample of 32, then the connections cannot be approved, so the test has to be expanded, and the final decision is delayed. A practical compromise is to test a sample large enough so that four connections could exceed 0.001 ohm slightly without causing m to exceed $m_{0.05}$. If $x = 1$ for the four connections, then $m = 4/N$ and, from equation 5, the corresponding sample size is 104.

175 C Screening Test

Although the 105 C test appears to be quite effective in distinguishing between stable connections and unstable connections, its slowness is a practical disadvantage. In cases where a number of alternate terminal designs are being considered, for example, and where it is desirable to concentrate development effort on a few of the most promising alternates, the 5-month waiting period can be

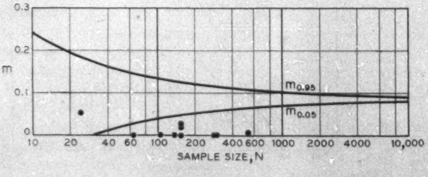


Fig. 10. Criteria for acceptance and rejection in 105 C accelerated aging test. Points are results of 105 C tests on approved types of solderless wrapped connections

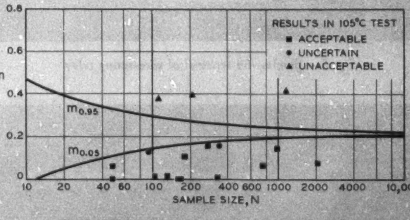


Fig. 11. Criteria for acceptance and rejection in 175 C screening test. Points are results of 175 C tests, but symbols indicate how similar connections behaved in 105 C tests

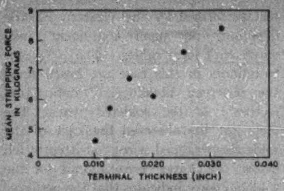


Fig. 12. Effect of terminal thickness on stripping force. Connections wrapped with no. 24 gage tinned copper wire on flat nickel-silver terminals 0.062-inch wide. Sample size 100

extremely inconvenient. There is need, therefore, for a quick screening test that will serve to identify those types of connections that are likely to pass the 105 C aging test.

Early experience with the 105 C test suggested that addition of mechanical disturbances to the original 3-hour undisturbed 175 C test might make it capable of detecting potentially unstable connections. Within limits, this proved to be so.

In the screening test that eventually was standardized, the solderless wrapped connections are baked in an oven for 3 hours at 175 C. The 3-hour period includes the warming up period of the oven and fixtures, which amounts to about one-half hour with the equipment that has been used. During the entire 3-hour period, each terminal is plucked automatically once per minute. The plucking mechanism deflects the free end of each terminal 1/16 inch, then releases it abruptly, allowing the terminal and lead wire to vibrate freely. The opposite end of the terminal is supported rigidly, of course, and the unsupported length (on which the connection is wrapped) is 19/32 inch.

At the end of the 175 C heat treatment, the connection is cooled to room temperature, and its resistance variation is measured as described previously except that 1. the terminal is plucked only once during the measurement instead of two or three times, and 2. it is plucked 1/16 inch instead of 1/32 inch.

In ordinary cases, where there is no soft solder in the connection, the correlation between the results of this test and the results of the 105 C test has been reasonably good. The principal discrepancy is that the 175 C test is not as sensitive as the 105 C test; that is, the spread between stable and unstable connections tends to be smaller in the 175 C test than in the 105 C test. This can be seen in Table I.

Despite its limitations, the 175 C screening test has been extremely useful in conserving testing effort and in speeding up various phases of the development program. Altogether, more than 16,000 solderless wrapped connections have been subjected to the 175 C test. On a few occasions, final approval of particular types of connections has been based upon results of the 175 C test, although the usual practice is to withhold final approval until the 105 C test is completed.

CRITERION FOR ACCEPTANCE (175 C SCREENING TEST)

Empirically, it is possible to define acceptance and rejection limits for the 175 C test that will correspond roughly to the limits for the 105 C test. Although such limits actually have not been used in the past, it appears that limits based upon $m'=0.22$ for the 175 C test would have led to essentially the same decisions that were made on the basis of 105 C test results. For $m'=0.22$, the limits for the 175 C test would be

$$m_{0.05}=0.22-0.77/\sqrt{N} \qquad (8)$$

and

$$m_{0.95}=0.22+0.77/\sqrt{N} \qquad (9)$$

These limits are shown in Fig. 11, along with observed values for various types of connections that have been subjected to both the 175 C and the 105 C tests. The symbols used for the values observed in the 175 C test indicate how similar types of connections fared in the 105 C test.

As indicated previously, the limits for the 175 C test can serve as guides for making decisions. If m for a particular type of connection is less than $m_{0.05}$, it probably is worth while to make tests at 105 C. If m is greater than $m_{0.95}$, it probably is not worth while. If m falls between $m_{0.05}$ and $m_{0.95}$, the decision will have to be based upon additional considerations.

The sample size for which $m_{0.05}=0$ is about 13. The sample size for which $m_{0.05}=4/N$ is about 41. A sample of 13, then, is the smallest sample that should be tested, and 41 is a more realistic minimum.

Mechanical Tests

Certain standard mechanical tests are made regularly to qualify wrapping bits for use in production, and the same tests have been made to qualify the bits used in the evaluation studies. Sample connections are wrapped on specified types of terminals, then two different types of tests are made on the sample connections. In the first test, the force required to strip the connection off of the terminal is measured. It is required that this stripping force be at least 3,000 grams and that the median for any subgroup of 5 connections be at least 4,200 grams. The purpose of this test is to provide assurance that the bit is capable of wrapping tight connections.

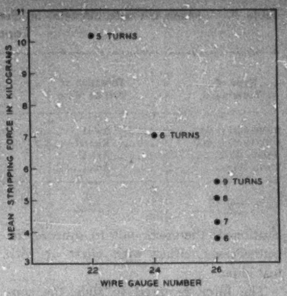


Fig. 13. Effect of wire size on stripping force. Connections wrapped with tinned copper wire on 0.0319 by 0.062-inch nickel-silver terminals. Sample size 50–175

In the second test, the wire is unwrapped from the terminal, the unwrapping force being applied to the wire at least one-half inch from the terminal without restraining the wire from twisting or bending back upon itself. It is required that the connection be capable of being unwrapped in this fashion without the wire breaking. The purpose of this test is to provide assurance that the bit does not wrap too tightly—so tightly that the wire would be weakened excessively.

Assuming that the connections have been wrapped with qualified bits, it is of interest to consider what might happen to them subsequently in service. In general, this reduces to a consideration of mechanical treatments that would tend to loosen the connection and that would tend to break the wire.

The principal types of mechanical treatment that could loosen a wrapped connection are 1. squeezing the sides of the connection, 2. pushing or pulling on the body of the connection, and, of course, 3. unwrapping the connection.

Although the squeezing forces that would be required to loosen connections have not been measured, it is evident that enough force could be exerted with a pair of pliers to damage a connection. Wiremen and maintenance men have been

Table II. Mean Ultimate Strength, in Grams, of Connections Made with No. 24 Gage Copper Wire

Type of Terminal	Direction of Pull on Wire	Solderless Wrapped Connections, 6 Turns	Wrapped and Soldered Connections, 1¼–2 Turns
0.0319 by 0.062	Axial	2,330	3,030
0.0319 by 0.062	Radial	4,814	5,205
Single Wire	Axial	4,450	4,444
Twin Wire	Axial	4,527	4,512

cautioned, therefore, not to squeeze the connections, either with pliers or with test clips.

The force required to slide the connection along the terminal ordinarily is well above the minimum requirement of 3,000 grams, ranging up to more than 10,000 grams in some cases. In general, the heavier terminals tend to give higher stripping forces than the lighter terminals, and the heavier gages of wire tend to give higher values than the lighter gages. Figs. 12 and 13 illustrate the relationships.

Ordinarily there is little risk that a connection will be unwrapped in normal wiring or maintenance operations—unless, of course, the wire is unwrapped deliberately to remove the connection. It is possible, nevertheless, to unwind a connection by pulling steadily on the lead wire in the direction parallel to, and in line with, the terminal axis. On the average, a force of about 825 grams is sufficient to unwrap one-half turn of a no. 24 gage connection on a 0.0319- by 0.062-inch terminal, and a force of about 2,300 grams will unwrap one full turn. The standard deviations are about 15% of these values.

The principal types of mechanical treatment that are liable to break wires in service are 1. tension alone, 2. repeated bending alone, and 3. repeated bending combined with tension. A number of tests have been made to compare the effects of these treatments on wires connected to terminals by solderless wrapping with the effects on wires connected by soldering.

The results of a few tensile tests are summarized in Table II. When the wire was pulled radially (perpendicular to the terminal axis), almost the full breaking strength of the wire was realized with the soldered connections. The breaking strength with the solderless wrapped connections was about 8% lower, however, because the wire was indented where it had been wrapped around the corners of the rectangular terminal.

It is interesting to note that, even with soldered connections, the full breaking strength of the wire was not realized when the wire was pulled axially. The soldered joint was broken in stages, and the wire was completely unwrapped from the terminal in most cases before the force reached the breaking strength of the wire. With the flat terminals, the ultimate strength of the solderless wrapped connections was significantly lower than that of the soldered connections. With the wire spring relay terminals, on the other hand, the differences between the solderless wrapped and the soldered connections were trivial. The serrations of the single wire terminal and the irregularities of the twisted and coined twin wire terminal were as effective as soldering in locking the wrapped wire in place.

The effects of repeated bending without tension are illustrated in Fig. 14. The test was performed by bending the wire back and forth through the angles indicated until the wire broke. The bending moments were large enough to exceed the elastic limit of the copper wire. In Fig. 14, conventional $\pm 3\sigma$ control limits (or confidence limits) are shown with each plotted point as a simple, graphic way of indicating that the performance of the solderless wrapped connections was significantly better, on the average, than that of the soldered connections.

Vibration tests, of course, provide another method for measuring the effects of repeated bending without tension. Fig. 15 shows the results of a vibration test performed by the Western Electric Company on small equipment units wired with local cables. The $\pm 3\sigma$ control limits in this case are based upon the observed breakage ("fraction defective") of the soldered wires. The fact that the observed breakage of the solderless wrapped wires consistently fell below the lower control limit for the soldered wires indicates that the performance of the solderless wrapped wires was superior on the average.

The effects of repeated bending combined with tension are shown in Fig. 16. In this test, the wire was kept under tension continuously while it was bent from its starting position perpendicular to the terminal axis to a second position parallel to the terminal axis and then back to its starting position. This cycle was repeated, always in the same direction from the starting position, until the wire broke. In Fig. 16, the $\pm 3\sigma$ control limits indicate again that the performance of the solderless wrapped wires was better, on the average, than the performance of the soldered wires.

From the results of tests such as those which have been described, it has been concluded that the solderless wrapped connection satisfies the mechanical security objectives in part, but not completely. It is more vulnerable to loosening by axial pull on the wire than is the soldered connection, so this limitation should be recognized in authorizing applications of solderless wrapping. In its ability to withstand vibration and repeated bending of the lead wire, however, the solderless wrapped connection appears to be fully as good as, and probably better than, the soldered connection.

Approved Combinations of Terminals and Wire

Present approvals of specific combinations of terminals and wire are based

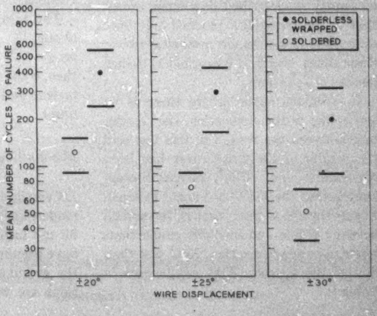


Fig. 14 (right). Effects of repeated bending without tension. Connections made with no. 24 gage tinned copper wire on single-wire terminals of wire spring relay. Sample size 16

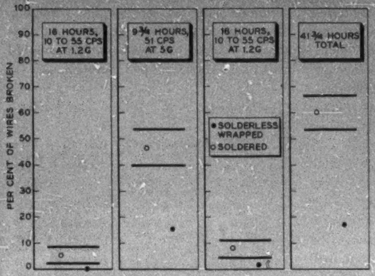


Fig. 15. Results of vibration test on small equipment unit wired with local cable. Sample size 550 for each type of connection

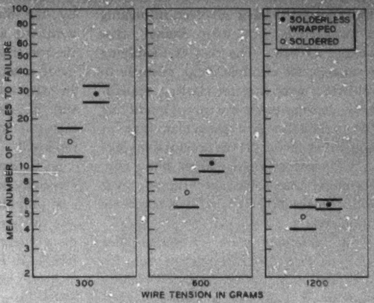


Fig. 16 (right). Effects of repeated bending with wire under tension. Connections made with no. 24 gage wire on single-wire terminals of wire-spring relay. Sample size 25 in each case

very largely upon the results of the accelerated aging tests that have been described, upon the ability of the manufacturer to manufacture the terminals economically, and upon a limiting dimension of a wrapping bit that is used widely.

The limiting dimension is the minimum diameter of the terminal hole in the bit used for wrapping no. 24 gage wire. In order to be sure that no. 24 gage wire can be wrapped on any approved terminal, the maximum dimensions of the cross section are limited so that the terminal will pass through a circular opening 0.073 inch in diameter.

Solderless wrapping with no. 22 gage tinned copper wire is approved on terminals of rectangular cross section whose nominal thickness is at least 0.030 inch and whose minimum diagonal exceeds 0.061 inch.

Solderless wrapping with no. 24 gage tinned copper wire is approved on 1. terminals of rectangular cross section whose nominal thickness is at least 0.025 inch and whose minimum diagonal exceeds 0.059 inch, 2. embossed terminals of the form shown in Fig. 6 punched and formed from flat stock whose nominal thickness is at least 0.010 but less than 0.025 inch, and 3. the wire-spring relay terminals shown in Figs. 7 and 8.

Approved terminal materials include nickel silver, brass, phosphor bronze, and silicon copper. The best terminal materials for solderless wrapping have a high modulus of elasticity, a low rate of stress relaxation, and a temperature coefficient of expansion near that of the wire.

In general, a copper flash plus an electrotinned finish is required on brass, phosphor bronze, and silicon-copper terminals, or they may be punched from either tin-coated or solder-coated stock. No finish is required on nickel-silver terminals, although any of the foregoing finishes is permissible and is specified in some cases to facilitate soldering. Solder dipping is not approved because of the risk of obtaining abnormally thick coatings from time to time and the undesirable effect that this could have upon the stability of connections wrapped on such terminals.

The use of untinned copper wire has been approved only in cases where the required service life of the connections is substantially less than 40 years. Furthermore, such approvals are limited to the heavy terminals which are approved for use with no. 22 gage wire.

There have been exceptions to some of the standards described in the preceding paragraphs. Solderless wrapping on solder dipped terminals and on thin, flat terminals was approved on a limited basis early in the development program. Those connections, however, are in circuits where trouble, if it should occur, would be detected automatically, could be located quickly, and could be corrected easily by soldering the defective connection. For future production, those terminals are being brought into agreement with present standards.

Performance in Service

The first field trial of solderless wrapped connections was installed in 1950, and limited use of solderless wrapping in regular manufacture began a few years later. During the period 1950–1958, the service performance of solderless wrapped connections appears to have been highly satisfactory.

A 2-year survey of 411,000 solderless wrapped connections in five central offices showed a lower wire breakage rate than would have been expected with soldered connections, and the difference was great enough to be considered statistically significant.

There has been no report of solderless connections being pulled off of terminals in service or of being partially unwrapped, and inspection of about 20,000 connections in two central offices revealed no sign of partial unwrapping.

The resistance variations of 135 connections were measured after 2 years in service, and 160 more were measured after $3^1/_2$ years of service. The highest value of ΔR observed was 0.0004 ohm, so the value of m for the 295 connections still was zero.

Conclusions

The laboratory studies and field experience to date have provided considerable assurance that properly designed and properly made solderless wrapped connections will perform satisfactorily for 40 years in central office service. The use of solderless wrapping is being expanded, therefore, in the Bell System. Suitable terminals are being provided now on many types of telephone apparatus. Design changes have been authorized to provide suitable terminals on a number of additional types of apparatus, although the changes have not been introduced yet in manufacture. And design changes to provide suitable terminals on still other types of apparatus are being studied.

Several hundred million solderless connections are being wrapped each year in the Bell System now. The number will grow, but it is difficult to predict what the saturation level will be. Inevitably, the solderless wrapped connection is in competition with soldered connections, clinched connections, welded connections, and all the other types. In the end, the choice of connections for any particular application is likely to be an economic choice, based not only upon the cost of the labor involved in making a connection, but upon many other factors as well. The cost of modifying terminals, for example, is an obstacle to solderless wrapping on existing apparatus. In some cases, the cost of modifying the terminals of a particular type of apparatus outweighs the potential savings of solderless wrapping. It is not likely, therefore, that solderless wrapping ever will displace other types of connections completely.

The present program for central office equipment in the Bell System calls for modification of the terminals of existing types of apparatus in those cases where the preparation expense clearly is outweighed by the direct savings from solderless wrapping and the indirect savings from the use of plastic insulated wire, which is made practicable by solderless wrapping. In other cases, modification of terminals will be deferred until present manufacturing tools wear out and have to be replaced.

In designing new types of apparatus, especially switching apparatus, the trend is to provide terminals suitable for solderless wrapping at the start. Furthermore, every effort is made to arrange the terminals in modular arrays that will facilitate automatic wiring by machines. These two steps are opening a door to economical use of automatic wiring in manufacture. It seems quite probable, therefore, that telephone switching systems of tomorrow will be well populated with solderless wrapped connections.

References

1. A New General-Purpose Relay for Telephone Switching Systems, Arthur C. Keller. *AIEE Transactions*, vol. 71, pt. I, 1952 (Jan. 1953 section), pp. 413–23; also *Bell System Technical Journal*, New York, N. Y., vol. 31, 1952, pp. 1023–67.

2. Solderless Wrapped Connections, J. W. McRae, R. F. Mallina, W. P. Mason, T. F. Osmer, R. H. Van Horn. *Bell System Technical Journal*, New York, N. Y., vol. 32, 1953, pp. 523–91.

3. Stress Systems in the Solderless Wrapped Connection and Their Permanence, W. P. Mason, O. L. Anderson. *Ibid.*, New York, N. Y., vol. 33, 1954, pp. 1093–1110.

Telex in New York

PHILIP R. EASTERLIN
MEMBER AIEE

AN AUTOMATIC dial teleprinter exchange service was placed in operation in May 1958, through which subscribers in New York City may obtain direct teleprinter connections, in a matter of seconds, to over 1,200 subscribers throughout Canada.

The new service, internationally known as "Telex," incorporates the characteristic automatic answer-back feature which confirms that the proper dial connection has been made and that transmission, even to unattended stations, is being correctly recorded on the distant teleprinter. The same subscriber equipment also can be used to connect to the land-line and cable tie-line sections for sending and receiving public message traffic.

The Telex Exchange at Western Union is linked by a number of carrier telegraph channels to the Montreal Exchange of the Canadian Telex network. The Canadian Telex service, comprising over 100,000 miles of telegraph trunk facilities, was placed in operation in July 1957 jointly by the Canadian Pacific and Canadian National Telegraph Companies, and its operation has been presented in an unpublished paper by Colombo.

Subscriber Set

In New York, the Telex subscriber set consists of a console-mounted type *15* teleprinter and an associated remote control (dial) unit shown in Fig. 1. A call is made simply by depressing a start button, dialing a five or six-digit subscriber number and the connection is made immediately after the last digit has been dialed. The fact that the connection has been made is indicated by the teleprinter motor starting and an operate lamp lighting. The connection can be terminated from either end by depressing a stop button which turns the equipment off. If the called subscriber is already engaged, a busy signal is returned which cuts the calling subscriber's equipment on for about one second, and then off again.

ANSWER-BACK FEATURE

The teleprinter keyboard in the subscriber set is equipped with an answer-back unit, a close-up of which is shown in Fig. 2. This device is a cylindrical drum associated with the keyboard which can be coded to transmit an answer-back sequence of 19 characters, 14 of which can be used to identify the

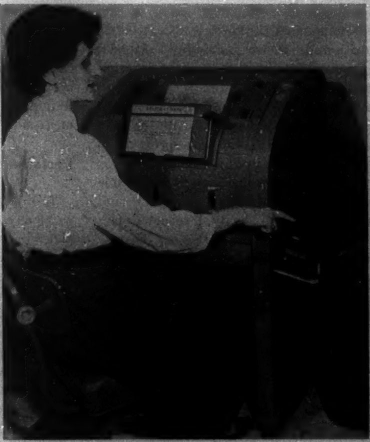


Fig. 1. Telex subscriber set

Paper 59-37, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 23, 1958; made available for printing November 21, 1958.

Philip R. Easterlin is with the Western Union Telegraph Company, New York, N. Y.

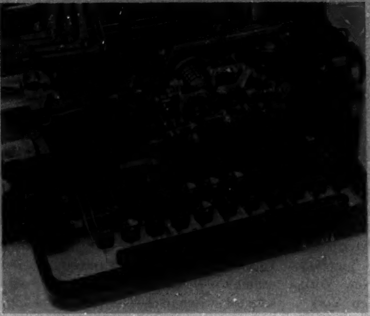


Fig. 2 (left). Answer - back unit installed on keyboard of type 15 teleprinter

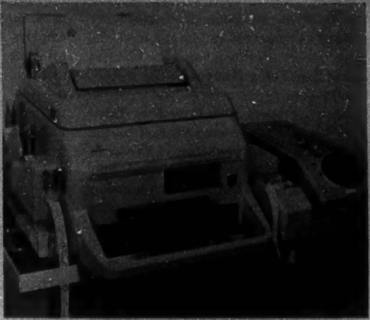


Fig. 3 (right). Subscriber set with local tape function

subscriber's name and city as listed in the Telex directory. The unit is remotely actuated by the WHO ARE YOU signal (FIGS. D) sent from the distant end, or it can be triggered locally by depressing an associated HERE IS key.

In normal operation, the calling subscriber initiates the exchange of answer-back sequences after a dial connection is established by transmitting the WHO ARE YOU signal, receiving the answer-back from the distant teleprinter and then depressing the HERE IS key to identify himself to the called subscriber. An exchange of answer-backs also may be initiated at any time during transmission to verify continuity in reception or as an over-all acknowledgment just prior to terminating a call. The answer-back, along with a paper-out protective circuit which prevents a connection being made to a subscriber's teleprinter having low paper, provides a high degree of reliability for Telex connections to unattended subscriber stations.

Subscriber Set with Local Tape Function

For Telex subscribers who require the additional operating facility of preparing, transmitting, or receiving on a 5-channel perforated tape, a special type of subscriber set is provided. This set, as shown in Fig. 3, is equipped with a Siemens model *100* page printer which has a built-in reperforator and transmitter-distributor attachment. A special type *L* remote-control unit is provided which allows the printer equipment to be used in an "off-line" manner for the advance preparation of tape. In the event of an incoming Telex call, the subscriber set is automatically switched "on-line" after a warning signal. Specifically, the set provides the following operation:

1. Off-line operation when idle, enabling the keyboard to be used for preparation of tape from reperforator attachment, along with a printed monitor copy.
2. After a connection is established, transmission may be from either the keyboard or by tape fed through the transmitter-distributor. Outbound, as well as inbound, transmission is again recorded on the page printer and the reperforator attachment may be used simultaneously to produce another tape.

The page printer used in this set also contains a 2-color ribbon feature which automatically prints the outgoing transmission from the keyboard and the incoming transmission in different colors. This is very useful when a 2-way Telex

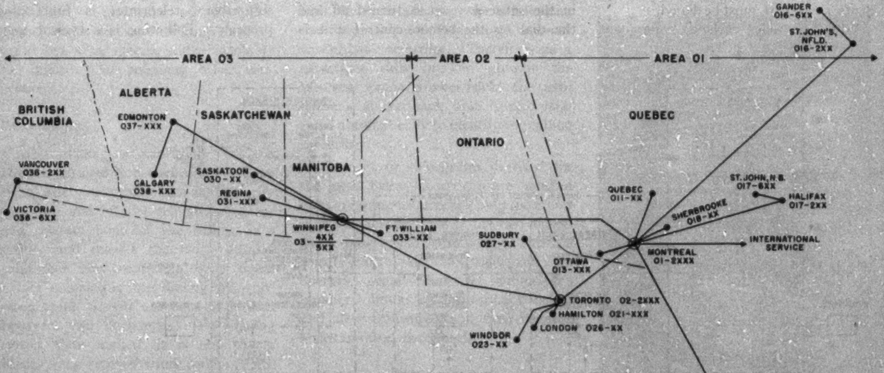

GANDER 016-6XX
ST. JOHN'S, NFLD. 016-2XX
AREA 03
AREA 02
AREA 01
BRITISH COLUMBIA
ALBERTA
SASKATCHEWAN
QUEBEC
ONTARIO
MANITOBA
EDMONTON 037-XXX
VANCOUVER 036-2XX
CALGARY 038-XXX
SASKATOON 030-XX
REGINA 031-XXX
VICTORIA 036-6XX
WINNIPEG 03-4XX/5XX
FT. WILLIAM 033-XX
SUDBURY 027-XX
OTTAWA 013-XXX
QUEBEC 011-XX
SHERBROOKE 018-XX
ST. JOHN, N.B. 017-6XX
HALIFAX 017-2XX
MONTREAL 01-2XXX
INTERNATIONAL SERVICE
TORONTO 02-2XXX
HAMILTON 021-XXX
LONDON 026-XX
WINDSOR 023-XX
NEW YORK 015-2XX

Fig. 4. Exchange layout of Canadian Telex network

conversation is later transcribed or read by other parties.

Canadian Telex Network

The layout of the Canadian Telex network and the interconnection to the New York Telex Exchange is shown in Fig. 4. The network is divided into three areas, centralized around a junction exchange in each area. These are area *01*, Montreal, area *02*, Toronto, and area *03*, Winnipeg. Other exchanges such as New York (*015*) are designated as district exchanges and are assigned three digits, the first two represent the junction area and the third digit denotes the district exchange within that area. The subexchange, which is usually in a nearby city to the district exchange, functions as an integral part of its respective district exchange by the assignment of a different fourth digit. For example, *036* designates the Vancouver, B. C., district exchange and *0366* switches a call through to a subexchange located at Victoria, B. C.

Local subscribers connected to the various junction and district exchanges are assigned either 5- or 6-digit dial numbers such as *03-564* for a Winnipeg subscriber or *015-222* for a New York subscriber. The digits on the right side of the hyphen denote the local number which is dialed to reach another subscriber who is terminated in the same junction or district exchange as the calling subscriber. For long-distance connections, calls to subscribers terminated by other junction or district exchanges, the complete number of digits, i. e., *03564*, must be dialed.

While the Telex exchange equipment used in Canada and New York City employs the direct dial or step-by-step principle and completes the connection as soon as the last digit has been dialed, it does impose certain restrictions on the routing of calls over trunk circuits. For example, the system does not normally employ registers for storing the dial digits and selecting alternate routes for calls when all trunks are busy or faulted. Therefore each long-distance call between two exchanges, excluding the subexchange, is given a predetermined route through one or more junction exchanges. For example, a New York call to Vancouver is routed New York–Montreal – Winnipeg – Vancouver and could not be passed through an additional Telex exchange such as Montreal–Toronto–Winnipeg, etc. Also, calls between local subscribers in subexchanges must be handled through the respective district exchange.

How a Telex Connection is Established

Before describing the technical functions of the switching equipment in the Telex exchange, a description will be given of the various steps which take place in establishing a long-distance Telex connection. A typical call would be where a New York subscriber dials *036-292* to reach a subscriber in Vancouver, B. C. The connection through the group-selector switches and repeater equipment in the various Telex exchanges for such a call is illustrated in simplified form in Fig. 5.

When not in use, the teleprinter motor in the subscriber set is turned off and the dial on the remote-control unit is locked. To initiate a call, the New York subscriber depresses a start button which causes the subscriber's preselector switch in the Telex exchange to seize an idle group-selector switch. Almost instantly, the exchange equipment returns a "go ahead dial" signal to the subscriber, unlocking the dial and bringing up a white visual indicator located in the remote-control unit.

The subscriber dials *036292*. While these digits are being dialed, the circuit is progressively built up through the various exchanges to Vancouver. For example, *0* causes the first group-selector switch in the New York Exchange to select a trunk to Montreal; *3* actuates the second group-selector switch in the Montreal Exchange to route the call to Winnipeg, and *6* extends the circuit in a similar manner from Winnipeg to Vancouver. The last three digits, *292*, are utilized in the Vancouver exchange in order to set the group selector and final selector switches to the respective subscriber.

Upon completion of the connection, requiring about 7 or 8 seconds, the teleprinter motor and a red operate lamp on the Vancouver subscriber set are turned on. As this occurs, a free-line signal is returned to the New York subscriber, also turning on the teleprinter motor and operate lamp to indicate that a connection has been established.

At this time an exchange of answer-back codes is initiated by the calling (New York) subscriber to insure that the proper automatic connection has been made and that the called (Vancouver) subscriber's teleprinter is functioning properly. Following is a typical oper-

Vol.
Pt. 1
195

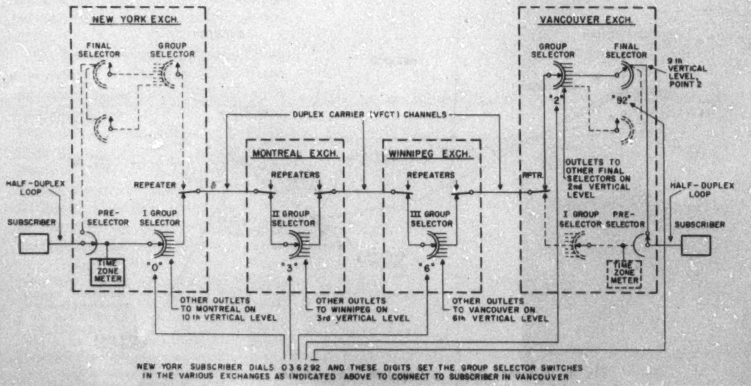

NEW YORK EXCH.
FINAL SELECTOR
GROUP SELECTOR
HALF-DUPLEX LOOP
SUBSCRIBER
PRE-SELECTOR
TIME ZONE METER
I GROUP SELECTOR
REPEATER
"0"
OTHER OUTLETS TO MONTREAL ON 10th VERTICAL LEVEL
DUPLEX CARRIER (VFCT) CHANNELS
MONTREAL EXCH.
REPEATERS
II GROUP SELECTOR
"3"
OTHER OUTLETS TO WINNIPEG ON 3rd VERTICAL LEVEL
WINNIPEG EXCH.
REPEATERS
III GROUP SELECTOR
"6"
OTHER OUTLETS TO VANCOUVER ON 6th VERTICAL LEVEL
VANCOUVER EXCH.
GROUP SELECTOR
FINAL SELECTOR
9th VERTICAL LEVEL, POINT 2
"2"
"92"
OUTLETS TO OTHER FINAL SELECTORS ON 2nd VERTICAL LEVEL
RPTR.
I GROUP SELECTOR
PRE-SELECTOR
TIME ZONE METER
HALF-DUPLEX LOOP
SUBSCRIBER
NEW YORK SUBSCRIBER DIALS 036292 AND THESE DIGITS SET THE GROUP SELECTOR SWITCHES IN THE VARIOUS EXCHANGES AS INDICATED ABOVE TO CONNECT TO SUBSCRIBER IN VANCOUVER

Fig. 5. Build-up of connection through several Telex exchanges

ating sequence which shows how the answer-back feature is used at the beginning and end of a Telex connection. Shown in parentheses is the control character which triggers the distant subscriber's answer-back unit and the symbols for the teleprinter non-printing functions, i.e., ← carriage return, ≡ line feed, ↑ figure shift, ↓ unshift.

Description of Operating Sequence	Characters Received by Both Teleprinters
WHO ARE YOU (Sent by calling subscriber)	(↑ D)
ANSWER-BACK (Sent automatically by called subscriber)	(←≡↓) SCOTIA BANK VCR (←≡)
HERE IS (Sent by calling subscriber)	(←≡↓) CHASE BANK NYK (←≡) (Telex message or 2-way teleprinter conversation)
WHO ARE YOU (Sent by calling subscriber prior to disconnecting)	(↑ D)
ACKNOWLEDGMENT (Sent automatically by called subscriber)	(←≡↓) SCOTIA BANK VCR (←≡)

Without respect to the originator of a Telex call, the subscriber at either end of the circuit may terminate the connection at any time by depressing the stop key on the remote-control unit. When this occurs, the trunk circuits through the several exchanges are released sequentially, beginning at the exchange of the originating or calling subscriber. This turns off the teleprinter motor and operate lamp on both subscriber sets, making them available for initiating or receiving other connections.

BUSY SIGNAL

After dialing a subscriber who is already engaged in a connection, a busy signal will be returned which momentarily turns on and off the teleprinter motor and operate lamp of the calling subscriber and relocks the dial. Also a subscriber may receive a busy signal before completing the dialing sequence when trunk circuits are busy or faulted between two exchanges. By dialing slower than normal, it is possible to identify the blocked trunk section. For example, if a busy signal is returned after the first three digits *036* have been dialed, it would indicate a shortage of trunk facilities between Winnipeg and Vancouver.

SUBSCRIBER SET OUT OF SERVICE

If a subscriber is dialed whose equipment or connection to the exchange is faulted, or where the paper roll is too low and needs replacing, neither a busy (which could be misleading) or a free-line signal is returned and therefore the subscriber set receives no response to the dial sequence. In such cases the calling subscriber is requested to depress the stop key, which cancels the call and recycles the dial unit, enabling the subscriber to re-dial and report the difficulty to supervisory personnel. If not cancelled in this manner, the trunk circuits involved in such a call are automatically released after a short period of time.

Public Messages Over Telex

The exchange equipment is arranged so that Telex subscribers can make free calls to the land-line or cable tie-line sections to send or receive public message traffic. As such connections are established, the subscriber or the telegraph company initiates an exchange of answer-back codes by following the regular Telex operating procedure. This speeds up the movement of message traffic by eliminating the delay in obtaining acknowledgments.

By dialing *1* or *3*, respectively, the New York subscribers are automatically connected to one of a group of page-receiving-only positions in the land-line tie-line section at New York, or to similar positions in the cable office. The receiving-only teleprinters are coded with the WHO ARE YOU signal following the Western Union Telegraph Company answer-back sequence. Therefore, as a subscriber calls in and obtains the Western Union answer-back, the subscriber's answer-back unit is also triggered automatically to identify the calling subscriber.

Messages or cablegrams destined for the local Telex subscribers are handled from a group of sending-only positions in the tie-line sections of the two offices. These positions are also equipped with a remote-control dial-box unit and the operator simply dials the subscriber's number, obtains an answer-back before and after transmission of each message series and then disconnects.

Information Center

An Information Center is located in the land-line tie-line section which New York subscribers can reach by dialing *9*. This Center communicates with other information centers in the Canadian exchanges to provide subscribers with current additions or changes in the Telex directory and to exercise general supervisory control over Telex network operations. Canadian exchanges may connect to the New York Information Center by dialing *015-9*.

Telex Charges

There is no minimum time charge for individual Telex calls between New York City and Canadian subscribers. Instead, there is a flat connection charge of 50¢ for each completed call, plus a particular pulse rate, based on airline mileage, which is applied for the actual time the connection is held intact. The pulse rate, having no relation to the telegraph transmission speed, is expressed in ppm (pulses per minute) and the charge for each pulse is $2^1/_2$¢. In addition, there is a $35.00 monthly charge for the subscriber equipment. There is no charge for calls to Western Union other than the regular rate for telegrams or cablegrams.

For example, a pulse rate of 15 ppm is applied for a New York City–Montreal call and 30 ppm for a New York City–Vancouver, B. C., call. Therefore, the connection and pulse-rate charge for a 1 minute New York City–Montreal call is 50¢+(15×$2^1/_2$¢) or $87^1/_2$¢ and for the New York City–Vancouver call it would be 50¢+(30×$2^1/_2$¢) or $1.25. If these calls were held for a 2-minute period,

Fig. 6. Connection (top) and pulse rate (bottom) counters for New York subscribers

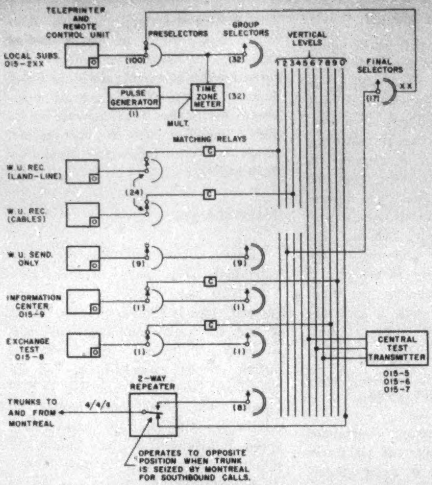

TELEPRINTER AND REMOTE CONTROL UNIT
PRESELECTORS
GROUP SELECTORS
VERTICAL LEVELS
1 2 3 4 5 6 7 8 9 0
FINAL SELECTORS
LOCAL SUBS. 015-2XX
(100)
(32)
XX
(17)
PULSE GENERATOR
(1)
TIME ZONE METER
(32)
MULT.
MATCHING RELAYS
C
W. U. REC. (LAND-LINE)
(24)
W. U. REC. (CABLES)
W. U. SEND. ONLY
(9)
(9)
INFORMATION CENTER 015-9
(1)
(1)
EXCHANGE TEST 015-8
(1)
(1)
CENTRAL TEST TRANSMITTER
015-5
015-6
015-7
2-WAY REPEATER
TRUNKS TO AND FROM MONTREAL
4/4/4
(8)
OPERATES TO OPPOSITE POSITION WHEN TRUNK IS SEIZED BY MONTREAL FOR SOUTHBOUND CALLS.

Fig. 7. Theory of New York Telex Exchange

Note: Figures in parentheses denote number of exchange units of each type

the charges would be $1.25 and $2.00, respectively. The connection and pulse-rate charges are assessed to the calling subscriber on two 5-digit counters individual to each subscriber line and are shown in Fig. 6.

Telex Exchange Operation

New York, as well as the exchanges in Canada, is equipped with a switching system which has been specially designed for teleprinter exchange service. While employing different circuitry, the system uses a number of equivalent telephone-exchange functional components such as preselectors, group selectors, final selectors, repeaters, etc. Fig. 7 shows a simplified theory layout of the New York Telex Exchange.

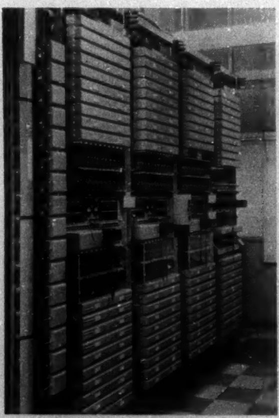


Fig. 8. Left to right, one rack for 20 final selectors, three racks of 40 local preselectors, and one rack of 20 long-distance preselectors

The operation of the exchange equipment will be covered by describing the functions which occur for setting up the typical New York City to Vancouver, B. C., call previously used in the brief explanation of how a Telex connection is established. The description for a northbound call, outgoing from New York and incoming at Vancouver, also will be typical of southbound calls from various Canadian exchanges to New York City. Photographs of the exchange-equipment units also will be referred to at appropriate points throughout the discussion. The multiple wiring and grouping arrangement of the equipment is not essential to a preliminary understanding of the exchange operation and is discussed later.

PRESELECTORS

Fig. 8 shows a group of preselector racks where the loop from each subscriber is terminated. Each subscriber's preselector consists of control equipment located in a subscriber bar and an associated preselector (stepping) switch capable of connecting to any one of ten GS/TZM (group selectors/time-zone meters). Each of the first three preselector racks accommodates 40 local subscribers (or tie-line positions) which are operated on a loop basis using make-break signaling. A different type of preselector rack, serving 20 long-distance subscribers, has additional control equipment for operating 2-wire polar over physical or carrier facilities for longer distances. In the latter case a d-c converter panel is required at the terminal of the distant subscriber to provide an equivalent termination between the line facilities and the conventional subscriber loop.

When the subscriber's set is idle, a current of about 5 ma (milliamperes) flows in the loop in a direction to keep the teleprinter motor cut off and the subscriber's remote control (dial) unit locked. As a subscriber initiates a call by depressing the start button on the remote control unit, equivalent to removing a telephone from the hook, the current in the loop is increased to 60 ma. This causes the subscriber's 11-point preselector switch in the exchange to leave its home position and seek connection to any one of ten GS/TZMs which may be idle. If all ten of these units are in use, the preselector switch steps to the 11th point, returns a busy signal to the subscriber and then restores to its home position.

GROUP SELECTOR/TIME-ZONE METER

The New York Exchange is equipped with 32 GS/TZMs and Fig. 9 shows five of these units per rack, with the group selectors located in the upper section and the five associated time-zone

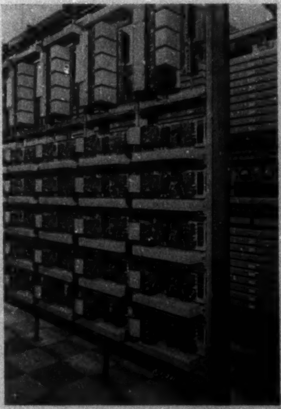


Fig. 9. Group selectors/time zone meters

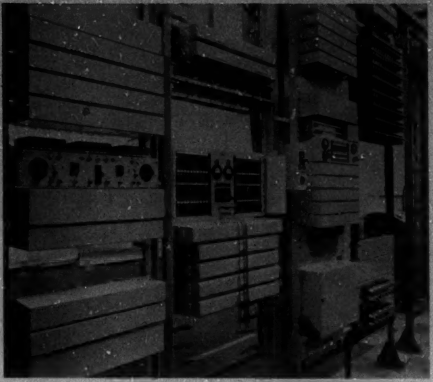


Fig. 10(left). Left to right, pulse generator, trunk repeater, and central test-transmitter racks

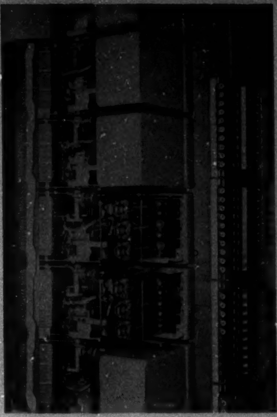


Fig. 11 (right). Group selectors associated with incoming trunk repeaters

meter units mounted below. The group selector is a 2-motion decade-type (stepping) switch with associated relay controls and the switch, having ten vertical levels (rows) and ten rotary points in each level, provides 100 outlets for three circuits designated as SEND, RECEIVE, and TEST, the latter being used for control purposes.

The associated time-zone meter is for the purpose of recording, by means of two stepping switches and associated relay controls, the first four dial digits, or less, in order to apply the proper charge for each call. In doing so, the proper metering pulse-rate conductor, i.e., 15, 18, 30 ppm, etc., is picked up from the pulse generator rack shown in Fig. 10 and is passed back over a separate conductor through the preselector switch to the subscriber's two individual connection and pulse-rate counters. After the dialing sequence and when a free-line signal has been given to the calling subscriber which indicates a completed call, both counters are stepped on a prepulse and thereafter only the pulse-rate counter responds to the metering pulses obtained from the pulse generator rack. The time-zone meter can be strapped to cancel and return a busy signal on certain calls which are to be rejected for special reasons. Also calls to the tie line or information positions at New York are recognized by this unit as "free" calls, for which neither counter is allowed to operate.

When the subscriber's preselector switch has made connection with an idle GS/TZM, which is almost simultaneous with the release of the start button, a go-ahead-dial signal is returned to the subscriber set. This signal is an interruption in the subscriber's loop current for about 25 milliseconds which unlocks the dial unit and brings up a white visual indicator in the remote-control unit, providing an indication that dialing can begin.

As the subscriber begins the dialing sequence, the group-selector switch in the GS/TZM is operated by the dial pulses of the first digit, which for a long distance call would be *0*, to the tenth level. The switch then free-hunts or self-steps horizontally over the ten rotary points on this level and stops on the first idle point corresponding to an outgoing or 2-way trunk repeater to the Montreal Exchange. A busy signal also will be returned to the subscriber at this point if all available trunk circuits are in use. Between the dialing time of the first and second digits, the trunk circuit to Montreal is tested by means of a revertive pulse which is described in the following paragraphs on the Telex trunk repeaters.

TELEX TRUNK REPEATERS

The 12 trunk circuits which connect to the Montreal Exchange are terminated in three types of repeaters designated as outgoing, 2-way and incoming, which are located on a rack shown in Fig. 10. The three types of repeaters permit the trunk circuits to Montreal to be operated on a 4/4/4 basis, which denotes the outgoing, 2-way, and incoming seizure priority for establishing a Telex connection. For example, the four outgoing trunk circuits may be seized only for northbound calls; the four 2-way circuits for calls in either direction, and the four incoming circuits for southbound calls only. A jack field and associated controls for testing individual trunk circuits is located in the center of the rack.

In addition to the switching controls for seizure of the trunk circuit, each repeater contains two polar telegraph relays which repeat the telegraph signals and dial pulses on a 2-wire polar basis between similar telegraph relays in the subscriber's preselector located in the exchange and the sending and receiving legs of the carrier channel terminal equipment. When a trunk circuit is unseized, the spacing (positive) polarity is applied in both directions.

Returning now to the typical northbound call, as the GS/TZM, which was previously described, stops on the point of an idle outgoing or 2-way repeater after the first digit has been dialed, the sending leg of the carrier trunk circuit to Montreal is reversed to the marking (negative) polarity. This seizure signal is received in the Montreal Exchange on an incoming or 2-way repeater having an associated group selector. If the trunk facility and group-selector equipment is in proper working order, the Montreal repeater returns a revertive pulse of approximately 25 milliseconds to the New York Exchange. Reception of this signal in New York momentarily operates the receiving relay in the Telex repeater to the marking polarity and then back to spacing, verifying circuit and equipment continuity for the reception of dial digits. Failure to properly test a trunk circuit in this manner instantly produces a busy signal to the calling subscriber, removes this particular trunk circuit from service, and operates an exchange alarm.

As the second and third digits are dialed by the New York subscriber in the typical *036-292* call to Vancouver, the second digit, *3*, sets the group selector at Montreal, as shown in Fig. 5, to the third vertical level where it free-hunts, seeking an idle outgoing or 2-way repeater to Winnipeg. When the connection is made to such a repeater, the trunk circuit between Montreal and Winnipeg likewise is tested by means of the revertive pulse as previously described. Following this the third digit, *6*, is used in a similar manner at Winnipeg to extend the call to Vancouver. The call now enters the Vancouver exchange in an incoming or 2-way repeater having an assigned group selector as shown in Fig. 11, which is pulsed by the fourth digit, *2*, to the second vertical level. Between the dialing time of the fourth and fifth digits, the group selector free-hunts over the rotary points in level *2* and connects to the first idle final selector.

For each 100 subscriber lines in an exchange a rack of 20 final selectors, as shown in Fig. 8, is provided for handling incoming calls to this group of subscribers. The final selectors also use the 2-motion decade-type 100-point stepping switch and associated relay controls, but unlike the group selectors, they do not free-hunt in a horizontal direction. Instead the final selectors are pulsed vertically and then horizontally by the remaining two dial digits to make connection to any one of 100 subscribers served by this rack.

As the group selector seizes an idle final selector and the dial digits *92* are received in the Vancouver exchange, the circuit is extended to the called subscriber's preselector switch. If the Vancouver subscriber is idle, as indicated by the preselector switch being in the home position, the relay controls of the preselector function to reverse the 5 ma flowing in the loop. This reversal of current actuates the polar motor-control equipment in the subscriber's set, which turns on the teleprinter motor, operate lamp, and also increases the loop current to the 60-ma working level.

Provided the loop current has been raised to the working level, the called subscriber's preselector in the Vancouver exchange will now return the free-line signal to the calling subscriber's preselector in the New York Exchange. This is a reversal of transmission polarity from spacing to marking in the east-bound direction. As this signal is received at New York the 60-ma current flowing in the calling subscriber's loop is reversed, causing the teleprinter motor and operate lamp to be turned on and the dial unit to be shunted from the loop circuit. As the connection is completed, the calling subscriber's connection counter is stepped one unit and the associated pulse counter begins recording the assigned pulse-rate charge for this call. Starting of the calling subscriber's teleprinter motor now permits the exchange of answer-back sequences previously described.

DISCONNECTING

Either the calling or called subscriber may terminate a Telex connection by depressing the STOP key on the remote-control unit for about 1 second. This transmits a steady-spacing polarity to the distant end, longer in length than normal teleprinter-signal combinations. Regardless from which terminal such a disconnect signal emanates, it is recognized by the calling subscriber's preselector, in this case New York, which in turn transmits a steady spacing signal through the GS/TZM, outgoing or 2-way repeater at New York, through the Montreal, Winnipeg, and Vancouver exchanges. This signal effects a release of the trunk circuits and switching equipment in the various Telex exchanges in the same sequence in which the connection was originally established. As the disconnect takes place, the subscriber sets at both ends of the circuit are restored to the idle condition by having the teleprinter motor and operate lamp turned off and the dial unit at the calling station relocked.

DIALING WESTERN UNION

Referring to Fig. 7, the local position preselectors are connected through matching relays, *C*, directly to specified levels on the GS/TZM switches, effecting a savings in final-selector equipment. As a New York subscriber dials *1*, *3*, etc., the GS/TZM switch is stepped vertically to the proper level where it then free-hunts and connects to the first idle position.

Other Exchange Equipment

The following paragraphs will briefly describe the operation and arrangement of other equipment in the exchange not explained in the foregoing description of a typical connection.

PULSE GENERATOR

The various pulse rates, one of which is picked up by the time-zone meter for each chargeable call to step the subscriber's pulse-rate counter, is supplied from the pulse generator rack shown in Fig. 10. This rack contains two pulse generators, one working and one fall-back, a changeover frame and 18 output relays, producing various outputs from 1 up to 36 ppm. In case of equipment failure, the companion pulse generator is automatically placed in operation.

The individual pulse generator consists of two polar telegraph relays having an R-C oscillatory circuit adjusted for a basic output frequency of 360 operations per minute. This output is used to drive several groups of switching relays, in a frequency-reducing network, which in turn drive 18 output relays corresponding to the required pulse rates. For the pulse and no-pulse intervals, a combination of negative battery and ground potential, respectively, is applied from the contacts of each output relay to the various pulse-rate leads which connect to the time-zone meters. The absence of this combination for any length of time enables an automatic pulse-control monitor to detect an open, short, or ground on any of the pulse leads.

In addition to the tariff function, a selected number of these pulse leads are used, instead of electronic timers, to actuate alarm and control circuits for abnormal exchange functions. For example, when a subscriber depresses the start button and does not dial a number, the exchange equipment is restored to normal after about 3 minutes by a control circuit which absorbs three of the 1-ppm pulses.

CENTRAL TEST TRANSMITTER

The New York Exchange is equipped with a central test transmitter, also shown in Fig. 10, which automatically transmits a choice of three types of tests for the use of maintenance personnel in checking local subscriber equipment or transmission over Telex trunk circuits. The characters for the test transmission are programed by relays from a drum-type distributor and each test can be transmitted simultaneously to five different lines.

When the digits *5*, *6*, or *7*, respectively, are dialed from a local subscriber set, the following tests are continuously provided by this transmitter:

1. A printing test utilizing all the lower and upper case teleprinter characters.
2. An unbiased "quick brown fox" sentence.
3. Five lines of "quick brown fox" in which one half of the sentence is alternately spacing and marking bias. Each line is prefixed with 0, 20, 30, 35, or 40% to indicate the amount of bias being transmitted.

Canadian exchanges may also obtain

these three tests automatically from New York by dialing *015-5*, *015-6*, and *015-7*, respectively, in order to check northbound trunk-circuit transmission.

EXCHANGE MULTIPLING AND GROUPING

The rack multipling between the stator points of all group and final-selector switches is made with 15-conductor ribbon wire which can be seen in Fig. 11. The ten points on the preselector switches are multipled in groups of 20 switches each which are wired to ten GS/TZMs and some of the latter are shared by several preselector groups. Ten group selectors are multipled in one group. Where the number of outlets exceed the ten points on the vertical level of these switches, such as levels *1* or *3* for the Western Union receiving positions or level *2* for the input to the final selectors, a mixing-plan arrangement is used and in some cases several groups of group selectors share the same outlet.

Life Characteristics of Carbon-Film Resistors after 12,000 Hours of Operation

H. BRANER
NONMEMBER AIEE

J. L. EASTERDAY
NONMEMBER AIEE

THE DEVELOPMENT of carbon-film resistors since World War II has provided the electronics industry with a reasonably priced semiprecision resistor.[1,2] The film resistors are smaller, lighter, cheaper, and have better frequency characteristics than precision wire-wound resistors. They are not as accurate as precision wire-wound resistors, and the temperature coefficient is still undergoing improvement, but the film resistor can perform an important function in filling the needs for a semiprecision resistor.[3]

Considerable effort has been devoted to determining the quality of these resistors, and minimum standards have been established, such as those found in *MIL-R-10509*.[4] However, little has been done to determine the long-term stability of carbon- and borocarbon-film resistors.

The purpose of a research program, conducted at Battelle Memorial Institute under the sponsorship of the U. S. Army Signal Research and Development Laboratory, Fort Monmouth, N. J., was to investigate the stability or drift of carbon-film resistors when operated at rated load in an ambient temperature of 70 C (degrees centigrade) for 12,000 hours.[5] Past work in this area has been confined primarily to the study of small sample sizes, various operating conditions, and relatively short periods of time such as 1,000 or 2,000 hours. Such work is generally done by the manufacturer to check his production, or for military qualification and acceptance tests. The 18-month program conducted by Battelle was designed with large sample sizes and carefully controlled test conditions so that statistical estimates of infinite population behavior were practical.

Laboratory Investigations

SAMPLES

The carbon-film resistor types investigated were 1/2-watt 237,000-ohm (critical values) resistors, and 1/2-watt borocarbon resistors with a nominal resistance value of 100,000 ohms. Four brands of resistors were included; three of these were deposited-carbon film and consisted of two sample groups of 250 resistors each and one sample group of 500 resistors. The fourth brand of resistors represented was a sample of 500 borocarbon resistors. Three of the brands investigated were supplied in a molded case, while the resistors of a fourth manufacturer were solder sealed in a ceramic tube.

TEST PARAMETERS

The resistors were operated at full rated load in an ambient temperature of 70 C for a period of 12,000 hours, with one exception. The borocarbon-resistor test was terminated after 9,500 hours of operation.

Resistance values were read after the resistors were installed in the chambers but were still at room temperature. Then heat was applied and the resistors were continuously monitored until the resistance values appeared to be stabilized. This required three to four hours. After the resistors were stabilized at 70 C, readings were taken which were used as the initial conditions against which all subsequent data were compared. After these readings were taken, the excitation voltage was applied in cycles of $1^1/_2$ hours on, and 1/2 hour off. At the end of the evaluation, the ambient temperature was reduced to room temperature and resistance deviations again recorded. The chambers were then opened for general observations and to determine the modes of failures.

EQUIPMENT AND INSTRUMENTATION

One of the important phases of this program was the design and construction of suitable temperature chambers. The factors considered in designing the temperature chambers were the effects of heat radiation from one resistor to another, air circulation, temperature gradient, and temperature control.

Optimum mounting distances between resistors were determined by preliminary tests. It was found that a horizontal spacing of 2 inches and a vertical spacing of 4 inches between resistors resulted in a maximum temperature rise of 0.3 C above ambient at a point midway between two resistors. Radiation baffles were provided by constructing the trays in the form of an open gridwork which resulted in a rectangular compartment for each resistor. Each chamber contained five of these trays with 100 resistors in each tray. Effects of heat interaction between resistors was further minimized by staggering the resistor positions in the vertical plane.

The additional heat required to maintain a 70 C ambient was introduced by

---

Paper 59-36, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted October 31, 1958; made available for printing December 1, 1958.

H. BRANER and J. L. EASTERDAY are with Battelle Memorial Institute, Columbus, Ohio.

The authors are grateful for the suggestions and information provided by H. Cary, W. E. Chapin, C. A. Shearer, E. N. Wyler, and the staff of Battelle Memorial Institute.

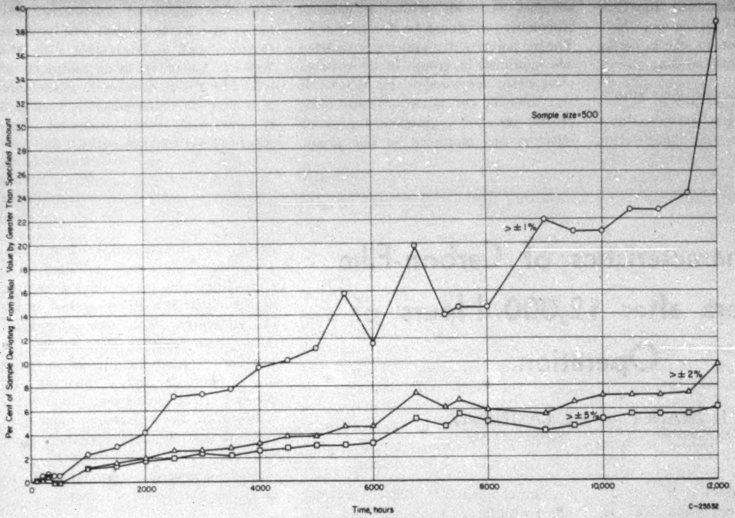


Fig. 1. Load-life characteristic, manufacturer A

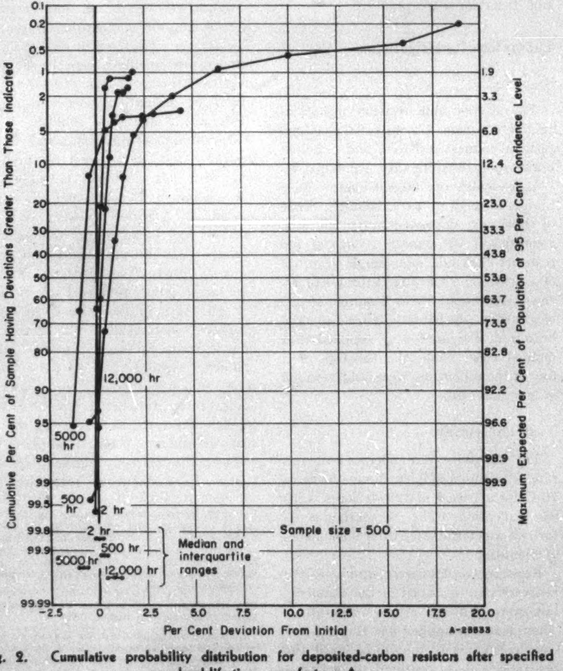
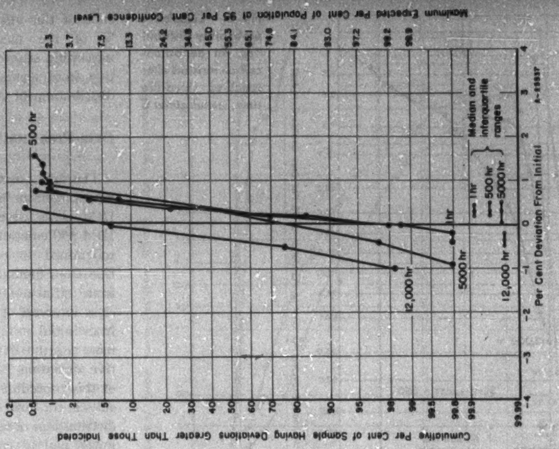


Fig. 2. Cumulative probability distribution for deposited-carbon resistors after specified load-life time, manufacturer A

an air flow, counter in direction and equal in magnitude to the natural convection currents, thus neutralizing the latter's effects.

Further effects of temperature irregularities were minimized by using a randomized distribution of samples throughout the oven. In addition, the load cycling was planned so that 75% of the resistors were energized at all times and that the energized resistors were always distributed throughout the chamber. The chamber temperature was maintained at 70 ± 5 C and the ovens had a gradient of 3 C. These temperatures were monitored with 15 thermocouples in each chamber, placed in groups of three for each tray.

The other important equipment problem was that of instrumentation. It was desirable to take the resistance readings of each group of 125 resistors during the 1/2 hour "off" portion of the energizing cycles, thus completing the data for the 500 specimens in a 2-hour period. To accomplish this aim, Battelle designed and built an automatic-scanning device which selects the components in a specified order and connects each component to a Barnes deviation bridge. The deviation bridge reads the per cent of resistance deviation from a standard (nominal resistance value in this case). The output of the deviation bridge is fed through a binary encoder to an IBM (International Business Machines Corporation) key punch. The key punch operates auto-



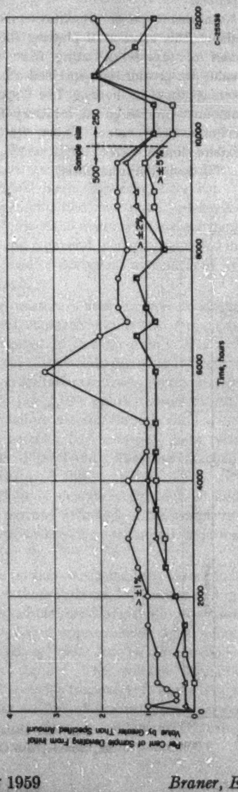


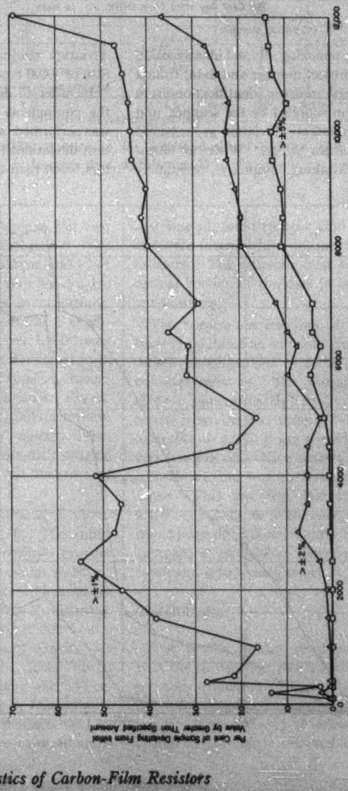
Fig. 3. Load-life characteristic, manufacturer B

Fig. 4 (right). Cumulative probability distribution for deposited-carbon resistors after specified load-life time, manufacturer B

500 specimens through 9,500 hours, 250 specimens after 9,500 hours

Fig. 5 (below). Load-life characteristic, manufacturer C

Sample size, 250 resistors



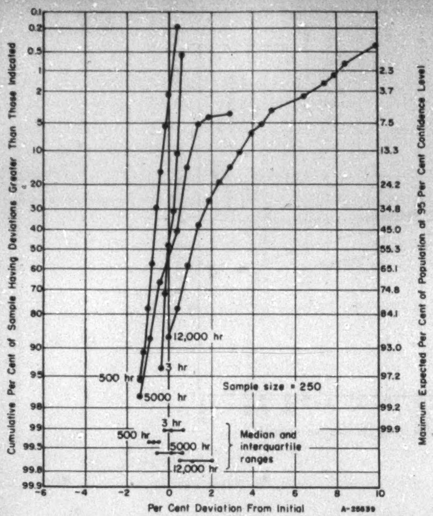


Fig. 6. Cumulative probability distribution for deposited-carbon resistors after specified load-life time, manufacturer C

matically, punching all fixed identification (such as project number and date) from a master card, resistor identification from information supplied by the scanner, and per-cent resistance deviation as measured by the Barnes bridge. With this automatic data-taking equipment, resistance-deviation readings can be taken at the rate of 1,400 readings per hour.

In order to determine the precision of the equipment, an analysis of variance was performed on 100 observations of per-cent deviations from nominal of ten resistors, each read ten times. From the results of this analysis we can expect 95 % of the measurements to have a measuring error of less than 0.5% deviation from nominal (assuming a normal distribution of errors).

Data Presentation

The data, showing per-cent resistance deviation from nominal, on the original IBM cards were processed through an IBM *650* computer. The computer was programed to convert per-cent deviation from nominal to per-cent deviation from initial and to group the results in class intervals of 0.5%. The deviations from initial were then arrayed from the most negative deviation to the most positive deviation. The ultimate objective of this procedure was to present the results in the form of cumulative frequency distributions of resistance deviations from initial plotted on arithmetic probability paper. A brief discussion of this type of graphic presentation is found in the Appendix.

Another type of presentation is used also. This consists of plotting the "per cent of sample deviating from initial value by greater-than-specified amount" versus time in hours. The "specified amounts" are, in effect, arbitrary failure criteria. The values chosen for these failure limits were ±1%, ±2%, and ±5% deviation from initial.

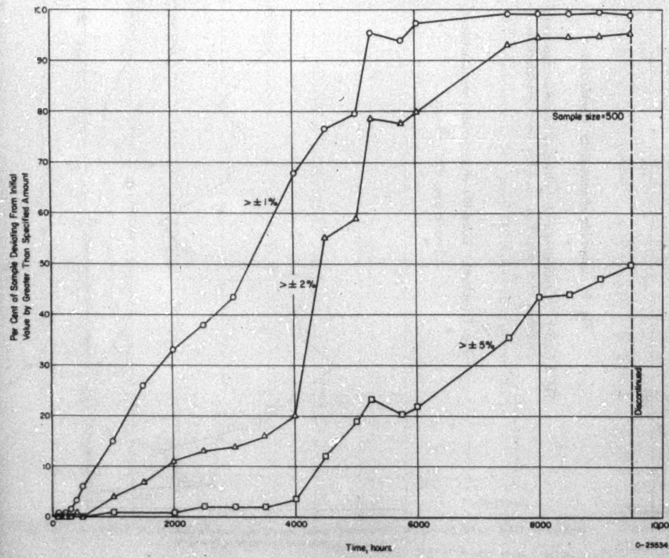


Fig. 7. Load-life characteristic, manufacturer D

Discussion of Results

The film resistors evaluated in this program displayed different aging characteristics depending on the manufacturer. The only generalizations to be made are that the resistors stabilize early in life and deteriorate at a constant rate until wear-out is reached.

Observing the ±2% deviation curve in Fig. 1, it is seen that the resistors of manufacturer *A* have a failure rate of approximately 0.7% per 1,000 hours until the 11,500-hour point. The change of failure rate at this point seems to indicate that a wear-out point has been reached. The erratic fluctuations noted in the curves are not necessarily indicative of erratic resistor behavior. As mentioned in the section on "Data Presentation," the percent deviations were divided into class intervals of 0.5%. Therefore, a large fluctuation on the curve may indicate that a quantity of resistors had a deviation change of 0.1 to 0.5% from the previous reading, thereby causing them to fall outside or just inside the fixed limit assigned to the particular curve. Cumulative distributions of resistance deviations at four representative times are shown for manufacturer *A* in Fig. 2. The linearity and the steepness of the slope of the three curves grouped about zero-per-cent deviation indicate tight distributions centered near zero for the first 5,000 hours of life. By 12,000 hours, however, 17 resistors had open-circuited, four more had deviations from initial which were greater than 30%, and a general increase was noted in the spread of deviations as evidenced by the graph.

The resistors of manufacturer *B* exhibit the best characteristic of the four types investigated in this program. After an apparent stabilization period of 100 hours, the average failure rate was approximately 0.13% per 1,000 hours (based on the ±2% failure criteria) through the 12,000-hour period. No wear-out point was noted. Figs. 3 and 4 illustrate the characteristics of these resistors. The distributions of deviations are normal and quite narrow, with 98% of the sample remaining within 2% of the initial deviation (see Fig. 3).

One mishap occurred during the evaluation of these resistors. As noted in Fig. 3, the sample size changed abruptly from 500 to 250 at approximately 9,000 hours. This change was due to an unreliable power supply. The voltage regulator failed in such a manner as to apply a high voltage to 250 of the resistors, thereby damaging them to an extent requiring their removal from the evaluation.

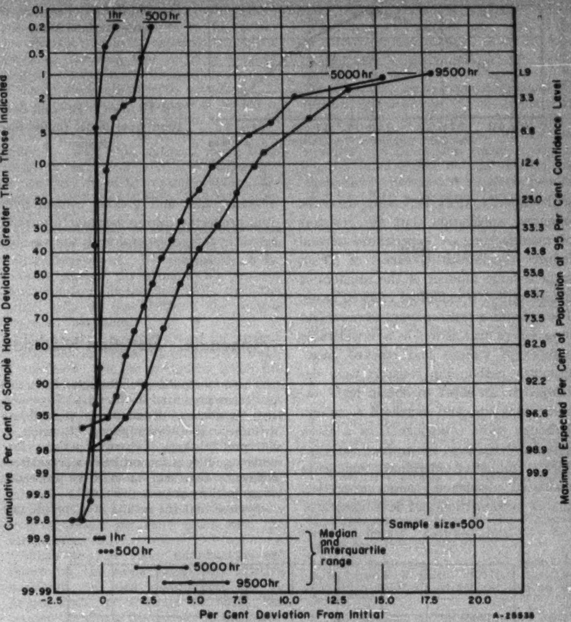


Fig. 8. Cumulative probability distribution for deposited-carbon resistors after specified load-life time, manufacturer D

From the remaining sample of 250, two of the resistors of manufacturer *B* open circuited during the evaluation, and 29 of them exhibited a series of very fine cracks, similar to crazing, on the outer surface.

The resistors of manufacturer *C* exhibited a failure rate of 2.3% per 1,000 hours from 500 hours to 11,500 hours. At 11,500 hours a wear-out type of characteristic is apparent, as noted in Fig. 5. A comparison of the 1% failure criteria for the different types of resistors shows that more of these resistors drift outside the 1% limit than do the resistors of manufacturers *A* and *B*.

The cumulative distributions of deviations are shown in Fig. 6. The interquartile ranges are somewhat broader than for the other carbon-film resistors. A large number of failures and excessive deviations are apparent from the distribution of the deviations at 12,000 hours: 29 open circuits, two short circuits, six resistors with deviations from initial greater than ±30% among the 500 resistors on test. Several case failures were also noted. These were characterized by cracks around the end caps on 63 resistors, one ruptured resistor, and evidence of severe internal heat on nearly all the resistors. The excessive heat caused the central portion of the case to turn white and powdery.

The borocarbon resistors of manufacturer *D* exhibited an inferior aging characteristic in comparison with the other carbon-film resistors. The average failure rate is approximately 10% per 1,000 hours, based on a criterion of deviations in excess of ±2% being considered failures (see Fig. 7). No wear-out point was in evidence, nor was a stabilization point noted. This test was discontinued at 9,500 hours because of the excessive number of large deviations as indicated by the cumulative probability distributions for 5,000 and 9,500 hours shown in Fig. 8.

Conclusions

The results of this program indicate that the carbon-film and borocarbon-film resistors evaluated in this program will not meet long-life high-stability requirements. The resistors of manufacturers *A* and *B* have fairly reasonable stability

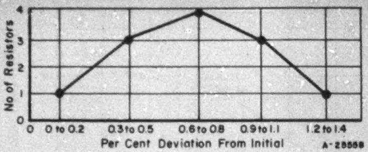
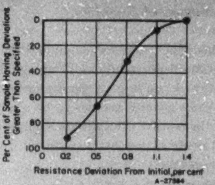


Fig. 9 (left). Per cent of deviation from initial

Fig. 10 (right). Per-cent deviation from initial (upper-class limit)



with respect to drift, depending upon the intended application, but the resistors of manufacturer *A* experienced several failures (deviations in excess of 30%), while a large number of the resistors of manufacturer *B* exhibited cracked cases. The remaining resistor types were subject to excessive drift as well as a considerable number of failures and cracked cases. Pyrolytic carbon-film resistors were developed in an effort to obtain high stability, long life, and increased accuracy. Although these components, as a class, were unsatisfactory, the authors feel that since a number of individual components evaluated displayed these properties, further investigation and development of resistor cases and greater control over film properties during manufacture could result in a fixed resistor with superior life characteristics.

Appendix. Statistical Methods

The major portion of the data from this project is presented in the form of cumulative frequency distributions plotted on arithmetic probability paper. In order to interpret this type of chart properly, a clear understanding of the nature of a cumulative frequency distribution and of the properties of probability paper is essential.

Assume that the results of a specific test performed on a sample of 12 resistors are as follows:

Resistor Number	Per-Cent Deviation from Initial
1	0.9
2	1.1
3	0.8
4	0.4
5	1.0
6	0.3
7	0.0
8	0.7
9	0.3
10	0.6
11	0.8
12	1.3

An ordinary frequency distribution constructed from these data might look like this:

Per-Cent Deviation from Initial	Number of Resistors
0–0.2	1
0.3–0.5	3
0.6–0.8	4
0.9–1.1	3
1.2–1.4	1
Total	12

In actual practice it would be impractical to construct a frequency distribution from so small a sample size. However, for the sake of the illustration, let us plot this distribution on ordinary arithmetic graph paper (Fig. 9).

This is the familiar normal distribution. If two additional columns are now added to the previous table, one showing the accumulated number of resistors having deviations from initial greater than the upper class limit, and the other showing this accumulated number as a per cent of the total number of resistors, the table would appear as follows:

Per-Cent Deviation from Initial	Number of Resistors, f_0	Number Greater Than Upper Limit, $n-f_0$	Per Cent of Total Greater Than Upper Limit, $\frac{n-f_0}{n}$
0–0.2	1	11	91.7
0.3–0.5	3	8	66.7
0.6–0.8	4	4	33.3
0.9–1.1	3	1	8.3
1.2–1.4	1	0	0
Total $n=12$			

The column showing the accumulated number of resistors greater than the upper-class limit is known as the "greater than" cumulative frequency distribution. The

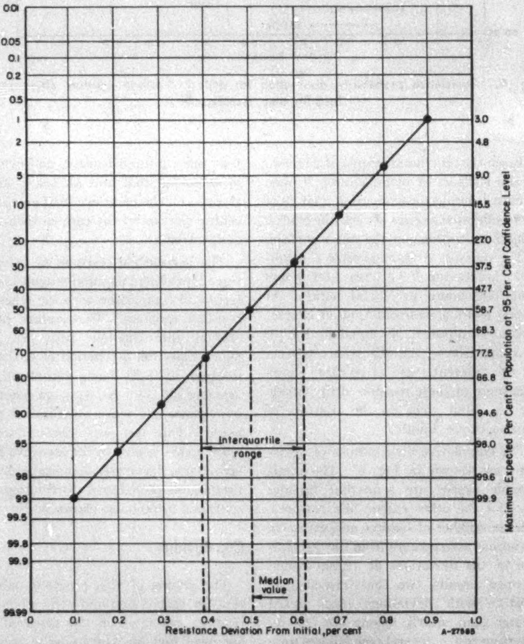


Fig. 11. Hypothetical cumulative distribution of resistance deviation from initial for 100 resistors after a specified test

column showing this cumulative frequency as a per cent of the total number of resistors is called the cumulative probability distribution because it represents an unbiased average estimate of the per cent of the population having deviations from initial greater than the upper class limit. If this distribution were plotted on arithmetic graph paper, an ogive would be described by the curve. This is shown in Fig. 10.

A special type of graph paper has been developed which is essentially arithmetic paper with the vertical axis marked off in per cent. However, since the cumulative probability curve of a perfectly normal distribution always yields an ogive of the same proportions, it was possible to adjust this per-cent scale by elongating the divisions in the middle and compressing the scale at the ends so that a normal ogive becomes a perfectly straight line. One other modification is necessary in the construction of probability paper. Statistical theory states that the entire range of population values cannot be estimated from a sample size which is smaller than the whole population. Therefore, the probability distribution, being a sample estimate of the population per cents, can never estimate the 100% or 0% point of the population. Thus the per-cent scale on arithmetic probability paper can never reach zero or 100%; consequently, the lowest value resistor and the highest value resistor cannot be shown on a cumulative probability distribution plotted on probability paper.

Fig. 11 is a graph of a hypothetical cumulative probability distribution of 100 resistors plotted on arithmetic probability paper.

A great variety of information is available from a graph such as this. Most apparent is the degree of normalcy of the sample distribution. This is evidenced by the "straightness" of the line, a straight line indicating a normally distributed sample. Any deviation from a perfectly straight line indicates an abnormally distributed sample, although in judging the linearity of the curve more weight must be given to the central portion of it than to the extremes. This is a result of the fact that in most random samples some extreme values will occur even if the population is perfectly normal.

A second type of information yielded is the resistance deviations within any chosen percentile range of the sample. Thus, for example, the 50% range (interquartile range) lies between the 25% and 75% points of the line, or, in the example shown, between 0.39% and 0.62% deviation from initial value.

If tolerance criteria are known, then the per cent of the sample exceeding tolerance can be determined. For example, if in our present example the tolerance for per-cent deviation from initial is 0.70%, then the graph tells us that 12% of the sample exceeded this value.

The graph can also be used for making population inferences at a specified confidence level. On the exemplary graph the maximum expected population per cent having deviation from initial greater than a specified value at the 95% confidence level is shown on the right-hand scale. Thus in our previous example where 12% of the resistors exceeded tolerance, the maximum expected population per cent exceeding tolerance at the 95% confidence level would be about 17.8%. The population estimate from a sample per cent varies with the sample size.

When two or more distributions are shown on the same graph, the relative variability of each distribution can be judged from the slope of the line. The greater the slope, the greater is the variability.

As mentioned before, the 0% and 100% points cannot be plotted on the graph. Hence the lowest value resistor in the sample of 100 and the highest value resistor are not shown, resulting in a curve which goes from 1% to 99% of the sample. On some of the charts in this paper it will be noted that more than one resistor is omitted from each end of the curve. When this occurs, it can be safely assumed that all, or all but two, of the resistors omitted had deviations from initial which exceeded the limits shown on the chart. If, for example, a graph of 100 resistors starts at the 3% of sample point and goes to only the 95% of the sample point, it must be assumed that at least four resistors had deviations greater than the highest value shown on the abscissa and that at least two resistors had deviations less than the lowest value shown on the abscissa.

References

1. PYROLYTIC CARBON RESISTORS: A COMPARISON OF THREE TYPES, Charles Wellard. *Electronic Equipment*, New York, N. Y., vol. 4, Apr. 1956, pp. 20–23; May 1956, pp. 28–30.

2. GLASS: A NEW LOOK IN RESISTORS, H. S. Craumer. *Ibid.*, vol. 3, Nov. 1955, pp. 16–17.

3. BORONCARBON RESISTOR CHARACTERISTICS, G. Kende. *Tele-Tech*, Philadelphia, Pa., vol. 11, Aug. 1952, pp. 48–49, 75.

4. *Military Standard MIL-R-10509.*

5. COMPONENT EVALUATION AND SPECIFICATION ENGINEERING, FINAL REPORT ON TASK 31, CARBON-FILM RESISTOR FAILURE RATES, C. A. Shearer, E. N. Wyler, H. Braner, J. L. Easterday. *Contract no. DA-36-039-SC-63136*, Battelle Memorial Institute, Columbus, Ohio, Oct. 15, 1957.

6. FIXED RESISTORS SHOW STABILITY IMPROVEMENTS, F. Rockett. *Electronics*, New York, N. Y., vol. 27, Sept. 1954, pp. 132–37.

7. FILM RESISTORS OF NITRIDE COMPOUNDS, E. H. Layer, E. R. Olson. *Electrical Manufacturing*, New York, N. Y., vol. 58, Sept. 1956, pp. 104–09, 310, 314.

8. HOT-MOLDING-KEY TO RESISTOR RELIABILITY, A. C. Pfister. *Electronic Equipment*, vol. 4, Oct. 1956, pp. 42–43.

9. RELIABLE PRECISION WIREWOUND RESISTOR DESIGN, J. S. Galbraith. *Transactions*, Professional Group on Component Parts, Institute of Radio Engineers, New York, N. Y., vol. CP-3, Dec. 1956, pp. 116–19.

10. MULTIPURPOSE EVAPORATED METAL FILM RESISTORS, S. J. Stein. *Ibid.*, pp. 119–24.

Rapid Determination of Corona Loss From Voltage-Charge Diagrams

H. S. DIXON
MEMBER AIEE

THE MEASUREMENT of a-c corona loss in laboratory test setups is frequently necessary when the magnitude of the loss is so small that wattmeter instruments cannot be used. An accurate, rapid means using only one measurement from a voltage-charge diagram is presented in this paper.

Voltage-charge diagrams were first used by Ryan to determine corona loss.[1] Time-consuming photography of oscilloscope patterns and planimetry thereof are required. From a laborious point-by-point analysis of corona voltage-current oscillograms, Karapetoff secured the corona loss.[2]

For magnetic samples, a method of determining hysteresis loss using three measurements from a hysteresis loop has been presented by Conover.[3]

Loss Determination Method

Following Ryan's voltage-charge method and using the basic circuit shown in Fig. 1, the diagrams shown in Fig. 2 were secured as the applied voltage was increased. Careful examination of a diagram such as that of Fig. 2(B) or 2(C) shows that the diagram consists essentially of a parallelogram plus two half-ellipses as shown in Fig. 3.

The area A of the complete diagram in terms of the quantities defined on Fig. 3 is

$$A=2E_iH+\frac{\pi H}{2}(E_m-E_i)\text{ volt-divisions}$$

This reduces to

Paper 59-24, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted March 31, 1958; made available for printing November 18, 1958.

H. S. DIXON is a Consulting Engineer, Berkeley, Calif.

The work described in this paper was performed at the High Voltage Engineering Laboratory, Purdue University, as a part of a Ph.D. thesis under the able guidance of Dr. C. S. Sprague.

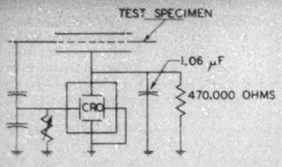


Fig. 1. Corona loss measurement circuit

$A = H(1.5708E_m + 0.4292E_i)$ volt-divisions

In terms of charge and energy A_e

$A_e = Q(1.5708E_m + 0.4292E_i)$ joules per cycle at the particular test frequency used

where Q is the charge in coulombs as determined from H. Consequently the energy loss per cycle at the test frequency is readily obtained from the applied voltage, the instantaneous corona initiation voltage, and a measurement of the vertical distance across the diagram when calibration of the vertical distance is made in coulombs. An example of such a calculation is given in the Appendix.

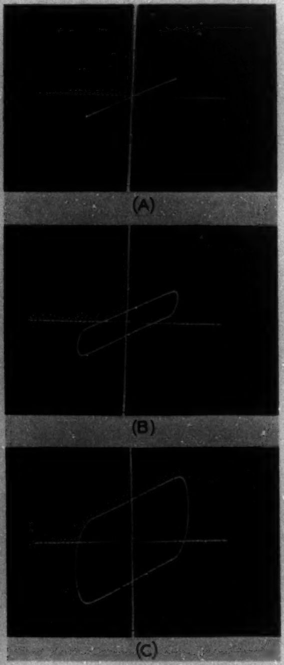


Fig. 2. Voltage-charge diagrams for three different voltages

A—35.2 kv, no corona
B—40.6 kv, corona present
C—47.2 kv, corona present

Unsymmetrical Diagrams

In the case of unsymmetrical diagrams such as shown in Fig. 4 which were secured with a positive d-c bias in series with the alternating applied voltage, the same equation for securing the a-c energy loss may be used. In this case H is measured so as to represent the average vertical distance between the two nearly straight line segments of the diagram. This procedure is equivalent to using the average of the major axes of the two half-ellipses.

Stabilization of Diagrams and Accuracy

Because of the extremely fast pulses or bursts during the negative half-cycle a-c corona, it is necessary to place a resistor across the integrating capacitor so as to stabilize the diagram on the oscilloscope so H can be measured. This resistor also drains off the charge that would otherwise accumulate on the capacitor as a result of the rectification action of corona.

This resistor also drains off some of the capacitor charge placed thereon during the corona bursts. Hence, the indicated value of H is reduced from its correct value by the presence of the resistor. However, if the time constant for the leakage of this charge is made sufficiently large compared to the half period of the applied voltage, the error is small and can readily be evaluated.

Assuming that during a half-cycle, the corona charge is placed on the capacitor by an instantaneous burst, it leaks off for an equivalent time of one half-period before the effect of this charge is measured by the distance H at the center of the diagram. Accordingly, the ratio of the charge measured q_m to actual corona half-cycle charge q_c is

$$\frac{q_m}{q_c} = e^{-\frac{t}{RC}}$$

where RC is the time constant and t is the half-period for the particular frequency used. For the apparatus used, the ratio q_m/q_c was 0.984, which means that H is 1.6% less than it would be if no stabilizing resistor were used. The measured power loss needs to be increased by 1.6%.

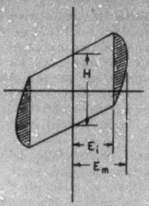


Fig. 3. Representation of area of symmetrical diagram by a parallelogram and two half-ellipses

E_i = corona instantaneous initiation voltage, volts
E_m = maximum instantaneous applied voltage, volts
H = average vertical distance across parallelogram, divisions

For the apparatus used, the capacitance between the cylinder and ground was calculated to be less than $1.0(10^{-4})$ microfarads so that its shunting effect on the 1.06-microfarad integrating capacitor was negligible.

To check the accuracy of the method of arbitrary representation of the corona voltage-charge diagram as a parallelogram plus two half-ellipses, the diagrams of Figs. 2 and 4 were planimetered. The calculated areas using measurements of H, E_i, and E_m from the diagrams were

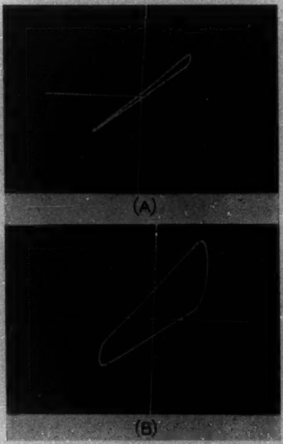


Fig. 4. Unsymmetrical voltage-charge diagrams

A—Triangular diagram
B—Quadrilateral diagram

the same as those secured by planimetry except for the largest area where the result was within the range of readings from individual traverses of the planimeter. Consequently, the method of calculating areas from the arbitrary representation is as accurate as planimetry, or is within 1%. Since this method of arbitrary representation of the diagrams is accurate to within 1%, this per cent error must be added to the per cent error in calculating A_e resulting from the combination of individual errors in measuring H, E_m, E_i, and the calibration data E_c, H_c, and C.

Conclusion

An accurate, rapid method of determination of corona loss from voltage-charge diagrams from one measurement on the diagram has been presented.

Appendix

Example of Calculation

The loss calculation for the following data, secured using a 1/8-inch-diameter rod centered in a 29-inch-long $6^1/_2$-inch-diameter cylinder, is given:

E_i=34,700 volts, corona instantaneous initiation voltage

E_m=35,500 volts, maximum instantaneous applied voltage

H=4.0 divisions, vertical distance across parallelogram

The data for the vertical calibration of the oscilloscope, and the integrating capacitor, where E_c is the rms voltage applied across the capacitor and H_c is the length of the vertical trace on the oscilloscope face, are

E_c=0.0190 volt

H_c=43.0 divisions

C=1.06 microfarads

Consequently

$Q=C2\sqrt{2}(E_c/H_c)H=1.06(10^{-6})(2\sqrt{2})(0.0190/43.0)(4.0)=0.0053$ microcoulomb per cycle

So

$A_e=0.0053[1.5708(35{,}500)+0.4292(34{,}700)]=375$ microjoules per cycle

The corona power loss at 60 cycles per second is

$P_c=60(375)(10^{-6})=23$ milliwatts

which is accurate to only two significant figures considering the accuracy of this particular value of H.

References

1. A POWER DIAGRAM INDICATOR FOR HIGH-TENSION CIRCUITS, Harris J. Ryan. *AIEE Transactions*, vol. 30, Apr. 1911, pp. 1089–1119.

2. CORONA ELLIPSES, Vladimir Karapetoff. *Ibid.*, vol. 48, Apr. 1929, pp. 317–26.

3. METHOD FOR THE DETERMINATION OF HYSTERESIS LOOP AREA, W. B. Conover. *Ibid.*, vol. 70, pt. II, 1951, pp. 1485–86.

A Talk-Back System for Educational Television

R. C. NORTON
NONMEMBER AIEE

J. B. DAVIDSON
NONMEMBER AIEE

TODAY the magic of television takes a teacher instantly to as many classrooms as desired. She appears before vast numbers of pupils not only with her voice, but with her material and her gestures, instructional ability, and personality intact. The principal missing element is the ability to know her pupils and to gage her pace by their reactions. Unseen and unheard, they lack presence.

In Cortland, N. Y., this major weakness of educational television is being overcome to a large degree by the talk-back system. This system, designed and provided by the New York Telephone Company, makes every pupil "present" to the studio teacher. By simply pushing a button she can instantly converse with any pupil in any classroom. Lamps on a panel before her indicate when the pupil has a question and stay on until she chooses to listen to his question. Thus hundreds of pupils at many and remote locations become integrated into one classroom, each pupil retaining his direct contact with the teacher and incidentally, with the other pupils since, with this system, all pupils hear the question or response from any classroom.

Although intended primarily for transmitting questions and answers between pupils and studio teacher it has been shown that the system, while not of program quality, is practical for the teaching of music classes with the studio teacher being able to listen to the singing of any classroom.

Operation

The system consists of two studios with associated control rooms and eight schools, each school having from two to six classrooms equipped for educational television. Initially 32 classrooms are in the system but the consoles and control equipment provided will serve an ultimate of 36 classrooms. Any classroom may be connected to either studio so that two classes may be conducted simultaneously. Fig. 1 shows a block diagram of the entire system.

Before a class is scheduled to begin the program director selects the classrooms to be used. The cabinet in the equipment room as shown in Figs. 2, 3, and 4 has a rotary switch for each classroom. He sets each to one of the following positions: OFF, STUDIO A, STUDIO B, or EQUIPMENT ROOM.

In this latter position he may listen directly to the particular classroom, all others being undisturbed. If he fails to reach the classroom because of failure of that teacher to turn on her power switch, he advances the rotary switch momentarily to a RING position. In this position a buzzer sounds in the classroom which is the teacher's signal to turn on her power switch.

As the teaching of the class proceeds the talk-back system remains normally "dead." If a pupil has a question, the local teacher in that classroom presses a push button on her desk. This lights a lamp on the console in the studio and also in the associated control room. These lamps remain on until the instructor acknowledges the question by verbally inviting the question and pressing the push button illuminated by the lamp. A red bulls-eye lights in the classroom as an "on the air" signal and the pupil asks the question without leaving his seat. Not only do those in the studio and control room hear the question, but, the occupants of all other classrooms hear it as well. Of course, by operating

---

Paper 59-229, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 17, 1958; made available for printing December 23, 1958.

R. C. NORTON and J. B. DAVIDSON are both with the New York Telephone Company, Syracuse, N. Y.

1. 78
1
59

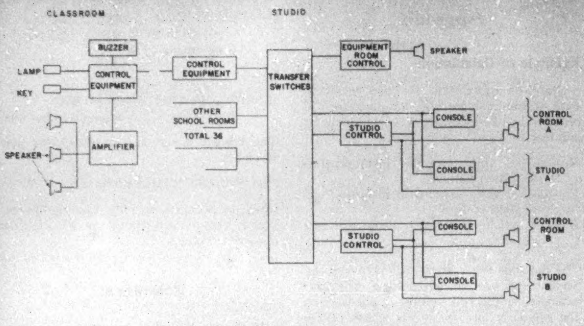


Fig. 1. Block diagram of talk-back system

the same push button the instructor can listen to the classroom at any time and receive answers to posed questions. An important feature of the system is that operation and control remain completely with the studio except for the initial signaling of a question which does not disturb either the studio or other classrooms.

Transmission Design

The principal design requirement of this system was that any pupil should be able to talk, without leaving his seat, to the instructor in the studio and to be heard by all other pupils in classrooms connected to the system.

The closed-circuit video facilities could not be used for this talk-back feature since direction of feed would have to be instantaneously reversed and problems of signaling and centralized control seemed insurmountable. Commercially available intercommunication systems were considered unsuitable for several reasons:

1. Some of the associated schools would be located many miles apart and hence should utilize telephone circuits with the attendant limitation of approximately 6 milliwatts of signal power.
2. The desired signaling, control, and intercommunication functions should, for economic reasons, be done with not more than a single pair of wires per classroom.
3. The duplicate signal lamp and control functions desired were not available.
4. Co-ordination with the audio portion of the television circuit was required, to avoid acoustic feedback during talk-back periods.

During the development of this new system tests were made using broadcast-quality microphones, telephone-type carbon microphones, high-quality loudspeakers, and small low-cost loudspeakers as microphones, singly and in multiple. The combination of weak voices, reverberant rooms, and high noise level made it difficult to achieve sufficient intelligibility. Under these conditions microphones were no better than small loudspeakers and, with the necessary preamplifiers, were much more costly. Loudspeakers and power amplifiers would still be required for the receiving condition.

The final design uses two small loudspeakers suspended overhead near the front of the room directed slightly downward toward the rear of the room. A third loudspeaker is placed on the teacher's desk. These loudspeakers, shown in Fig. 5, serve as both microphones and loudspeakers using a reversible amplifier which was designed for the "School-to-Home" service whereby a disabled child at home listens and talks to his classroom.

Contrary to the popular quest for high fidelity and flat frequency response this amplifier has a response designed to peak in mid-band for maximum intelligibility under classroom conditions. The frequency characteristic of the amplifier is shown in Fig. 6 which also shows, for

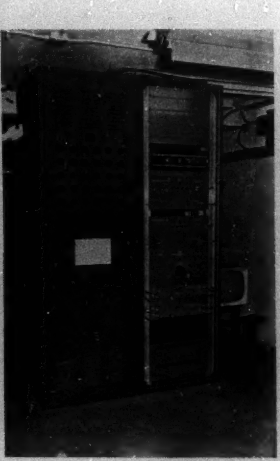


Fig. 2. Talk-back equipment cabinet on left and television equipment cabinet on right

Fig. 3. Control equipment and programing selector switches

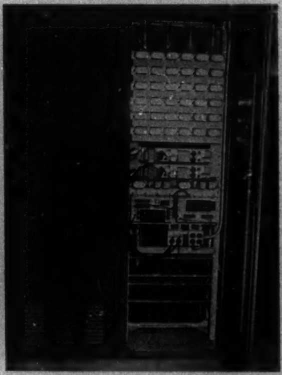


Fig. 4. Apparatus arrangement in equipment cabinet

Fig. 5. Installation arrangement in a typical classroom

comparison, the frequency characteristic of the common telephone which has evolved through the years as the best compromise between fidelity, intelligibility, and economics. The pass band of regular telephone message circuits is entirely adequate for this talk-back service thus avoiding the use of more costly "program" quality circuits such as are required for the best transmission of music.

Conventional intercommunication systems which use small loudspeakers as microphones are intended to provide 2-way talking circuits between two locations but usually one party or the other is close to his loudspeaker. The condition of a distant and, therefore, weak voice at one end attempting to talk to a person who is also distant from his loudspeaker is avoided automatically by the usual requirement that one party or the other "push-to-talk." The educational television talk-back system faces this problem in that a student in a distant corner of one classroom must be heard by a student in the distant corner of another classroom. The generally bare walls and floor which make the acoustics poor, the shuffling of papers, and the restlessness of school-age children all combine to make the transmission problem difficult. Multiple source origination, while generally to be avoided, becomes helpful under these conditions and since the same units are used as reproducers, a fairly uniform sound level is achieved throughout the room. A compressor-amplifier of the type frequently used in the radiobroadcast field assists greatly in overcoming the wide variation in volume between different children's voices and at various distances from the microphones. Levels are set with the amplifier operating in the middle of its range so that weak signals are expanded while strong signals are compressed, thus achieving automatic volume control.

The compressor-amplifier serves a dual purpose in that it is used to overcome the bridging loss of the many connected classrooms. A pad in its output circuit not only serves to make the level fed to the classrooms independent of the number of rooms connected, but also isolates each leg so that a trouble on any circuit will not disable the entire system. These pads can be so selected as to equalize the levels reaching schools located at various distances from the central bridging point. The pads affect only the outgoing level and differences in the incoming direction are absorbed by the compressor-amplifier.

At the school amplifiers normal sending levels range from +8 volume units for strong talkers down to the room noise level of −30 volume units. Within the city of Cortland the cable circuit loss ranges from 4 decibels for the nearest school to 10 decibels for the most distant. Transmission over a circuit longer than about 25 miles of loaded 19-gage cable would require telephone repeaters with due consideration being given to the simplexed control circuit. The relays in the control circuit impose a limit of 500 ohms for the simplex path or 2,000 ohms of conductor loop resistance which is equivalent to 23 miles of 19-gage loaded cable.

Equipment Design

In providing the equipment required for control, signaling, and switching many unusual design problems were encountered.

At the studio location the need was apparent for a console designed for complete simplicity of operation by the studio teacher or personnel in the control room. The desired features were instantaneous recognition of the schoolroom signal and immediate association with its control key, ease of operation, portability, appearance, and expansion. Various types of standard turrets and consoles were considered, but were found to lack one or more of the mechanical or other features desired. Three standard Western Electric 700-type turret cases as shown in Figs. 7 and 8 were combined for the console housing providing a good appearance and permitting ease of expansion since additional cases may be added as required. The illuminated type of push button normally used in the 6-button telephone set was found to be most desirable. These buttons embody the feature of direct illumination of the push button by an associated lamp and designation strips are provided as a part of the unit. A separate switch on the console permits talking back to the classrooms over this system if desired. A semiflexible cable was provided to facilitate movement about the studio.

In the equipment room a cabinet houses all the equipment required at the studio end. On the outside of the door affording desirable access to both the

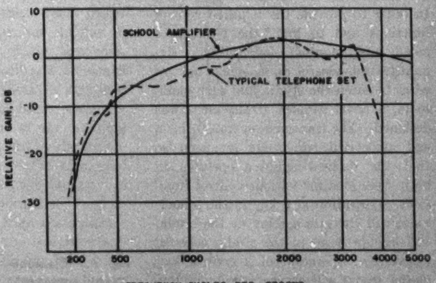


Fig. 6. Frequency response of classroom amplifier

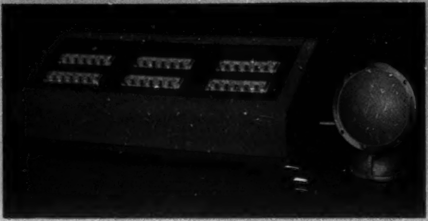


Fig. 7. Studio console

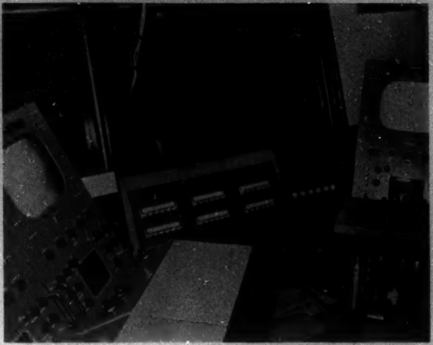


Fig. 8 (right). Control room console

program director and maintenance personnel, are mounted rotary-type switches. These switches connect the schoolroom circuits to the desired studio and provide required functions. Operational instructions and designations are also mounted on the door.

The low-voltage power requirements for this talk-back equipment at the studio are met by the new standard Western Electric *101-J* power plant. The console lamps use a 10-volt a-c supply with a maximum drain of 3 amperes and the relays require a 24-volt d-c supply with a maximum drain of 1.5 amperes. These requirements and provision for 50% growth are met by this power plant.

In the schoolroom all equipment was prewired and mounted on a backboard and then installed as a single unit as shown in Fig. 9. Power for the lamp and relays is provided by another new standard Western Electric power plant, the *101-H*. This is a small power plant which was designed for small key telephone systems. The output of this power plant is adequate for the schoolroom requirements.

The number of cable conductors required between the studio and classroom locations was of major importance. In order to provide the required control functions and to limit the number of conductors required for talking, signaling, and control to a single pair of wires between the studio and each classroom, a novel simplex arrangement was designed. The transmission transformers and associated relays are arranged so that the derived simplex operates in both directions for various control functions. Operation of a key in the schoolroom will energize a relay at the studio and light a lamp in the studio console. Conversely, operation of a control key in the studio will arrange the amplifiers at the studio and in the classroom for transmission and, in addition, light an "On the Air" lamp in the classroom.

The general over-all design features the use of standard Bell System equipment wherever practical. The maintenance personnel are familiar with this equipment and it is readily available from stock. Testing procedures will not require any special tools or equipment. Both of these items should expedite maintenance on the over-all system.

Selection and Operation

Selection of the classrooms to be associated with either of the studios is made by the program director at the equipment cabinet. Each schoolroom through its line circuit is associated with a rotary switch, *S2*, as shown in Fig. 10, that enables the schoolroom to be connected to either studio, the equipment room, or to be left disconnected from all locations. All classrooms to be associated with the studio program are selected prior to the start of program, according to the scheduled arrangement.

The teachers in the classrooms will operate the power switch associated with their equipment at program time. The operation of the power switch energizes the power plant which operates relay *B* disconnecting the schoolroom buzzer from the line. The power switch also energizes the schoolroom amplifier, and arranges its control equipment to function under control of the studio console over a simplex of the line circuit. The signaling key in the classroom will also operate over the simplex in the reverse direction. The speakers are normally on the output side of the schoolroom amplifier in the receive position.

Operation of the classroom rotary switch will connect its line circuit through to the output side of the bridging amplifier associated with the particular studio.

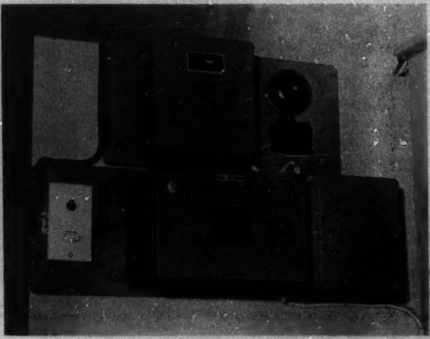


Fig. 9. Classroom control equipment

Vol.
Pt. 1
195

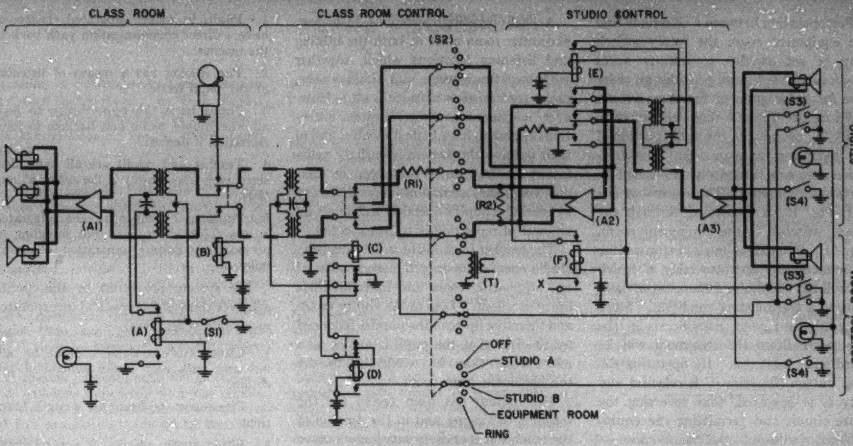


Fig. 10. System circuit

The rotary switch also connects the lamp and control leads of the schoolroom to the selected studio control consoles located in the studio and its control room. All schoolrooms associated with the particular program are selected in the same manner and consequently are all in the normal or receive position. The studio speakers are normally on the output terminals of the studio amplifier.

The studio line circuit is bridged to the schoolroom line circuits at the output terminals of the bridging amplifier. All stations therefore are normally in the listening position. A shunting arrangement is provided across the output of the bridging amplifier to prevent the classrooms from interrupting the studio during the silent periods. A muting arrangement on the television microphone is provided during the operate period to prevent acoustic feedback between the audio portion of the video network and the talk-back circuit.

During the progress of the lesson any associated schoolroom may indicate to the studio teacher that they wish to ask a question by momentarily operating the push button located on the teacher's desk. Operation of the push button will operate the *D* relay over the simplex of the line. The *D* relay will lock operated under control of its associated *C* relay and light its line lamp in the consoles in the studio and associated control room. The *A* relay in the schoolroom is also operated but performs no function at this time, transmission being prevented by the shunt on the output of the bridging amplifier. The schoolroom now awaits instruction from the studio teacher as the talk-back is completely under control of the studio. At the convenience of the studio teacher the schoolroom will be instructed over the television system to proceed with their question. The illuminated button in either the studio or control room console must now be operated, energizing the *C* relay which removes the shunt from the output terminal of the bridging amplifier and mutes the studio microphone. The operation of the *C* relay transfers the schoolroom line from the output to the input terminals of the bridging amplifier, removes the locking path to the *D* relay, extinguishes the line lamp, and grounds the simplex of the line circuit to operate the *A* relay in the schoolroom. The operation of the *A* relay lights the "On the Air" lamp in the schoolroom, and reverses the school amplifier which places the speakers on the input and the line on the output terminals of the amplifier. The classroom voices will now be picked up by the schoolroom speakers, amplified in the schoolroom amplifier, sent over the line to the imput of the bridging amplifier, and will be heard by the studio, control room, and all of the connected classrooms. Release of the button will restore the circuit to normal.

Audio transmission from the studio is normally carried over the television system; however, provision has been made that enables the studio teacher to talk over the talk-back circuit. A studio control key, *S4*, is provided at the consoles for this feature. Operation of the *S4* key operates the *E* relay which transfers the studio line circuit from the output to the input terminals of the *A2* amplifier. The *E* relay operates the *F* relay which removes the shunt from the output terminals of amplifier *A2* and mutes the studio microphone. The *E* relay also reverses the *A3* amplifier and places the speakers on the input and the bridging amplifier *A2* on the output terminals of the *A3* amplifier. The studio or control room may now talk to all connected classrooms.

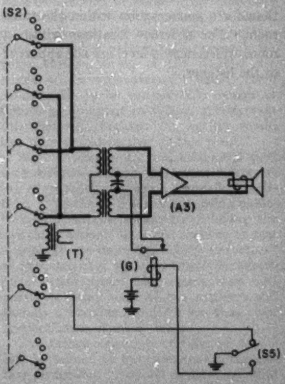


Fig. 11. Equipment room talk-back circuit

For operation between a classroom and the equipment room the rotary switch may be set at the EQUIPMENT ROOM position. This feature provides an order wire for preprogram or maintenance use and a means of monitoring a classroom to verify, by listening to the audio portion of the television program, that the television sets are turned on and tuned to the proper channel. The classroom circuit, Fig. 10, is connected through the rotary switches to the equipment room, Fig. 11. Ground at push-button switch *S5* immediately operates relay *C* which grounds the simplex and changes the classroom to a transmitting condition. Since the equipment room amplifier is in the receive condition the classroom will be heard and monitored. By operating the push button *S5*, relay *C* is released and relay *G* is operated, thus reversing the entire circuit and permitting the equipment room to talk to the classroom.

If a classroom teacher has failed to turn on her power switch, the studio will be unable to hear room noise from that room.

With the power switch off, relay *B* will be released, connecting the line to a buzzer. In the equipment room, operation of the *S2* rotary switch to the RING position will energize the buzzer and indicate to the classroom teacher that she should operate the power switch in her room.

Performance

Under test conditions a teacher can walk anywhere in the classroom and talk back and forth with a teacher anywhere in another classroom. When talking at a normal conversational voice level, intelligibility is satisfactory with requests for repetition being no more frequent than for a conversation within the same room. For different locations within a room, differences in level are not apparent to the listener.

Actual operating conditions, of course, encounter room noise at both the talking and listening locations which, together with sometimes weak and timid voices, make transmission more difficult. However, teachers accustomed to such conditions appear to have little difficulty, and in their opinion the system is entirely satisfactory. They know that even in their own room they sometimes have difficulty in hearing a pupil's question and a certain amount of repetition is to be expected.

The teacher back in the quiet television studio encounters only the room noise of the originating room and finds intelligibility excellent. Loss of the ability to see and therefore lip-read the pupil is balanced by the fact that the pupil is nearer to a microphone than he would be to the teacher in the same room.

The system has been tested for the teaching of singing and in the opinion of the teacher it is entirely satisfactory since she can instantly listen to any class and judge its pitch, timing, and performance of part singing.

The ability to switch from one room to another by the push of a button makes apparent the value of acoustic tile ceilings in reducing reverberations in the classrooms. About half of the classrooms have such ceilings and the improvement in intelligibility and naturalness from these rooms is very evident. Of course the acoustic ceiling is of great value to the teacher who teaches every day in that room and also improves the reproduction of recorded music used in music appreciation courses. The proof is most strikingly shown by the quick switch from classroom to classroom of the talk-back system.

Conclusions

This talk-back system is a valuable adjunct to an educational television system for several reasons:

1. Pupils in widely separated classrooms have a direct communication path back to the teacher.
2. The teacher has a means of listening to any pupil recite.
3. Pupils have the opportunity to hear all other pupils recite and partake in joint recitation if desired.
4. Teacher and pupils are all kept alert since each may question the other at any time.
5. Pupils and teacher become integrated into a joint effort with full mingling of personalities despite separation of many miles.

The system presented in this paper achieves these objectives and incorporates many salient features:

1. Classrooms can signal a desire to talk but may talk only under control of the studio.
2. Classrooms perform no control functions.
3. Either studio teacher or control room operator may assume control.
4. Individual classrooms may be split off from the system at any time for direct communication and for program readiness checks.
5. If classroom power switch is inadvertently in the OFF position the room may be called by a buzzer.
6. Only a single pair of wires is required from each classroom back to the studio.
7. Regular telephone circuits are used.
8. All components are standard unmodified items easily obtained if replacement is required.
9. The system can be expanded to care for additional schools and classrooms.

The most important criterion of any design is customer satisfaction. It need only be said that an experienced and critical customer is not only satisfied, but proud of the system just described, and believes that another step forward has been made in the quest for better methods of educating the youth of the nation.

Vol.
Pt. 1
195

LATE DISCUSSIONS

Discussion of "Organic-Moderated Reactors for Central Station Power," by W. E. Parkins and E. F. Weisner, published in Communication and Electronics, January 1959, pages 985–93.

A. C. Werden, Jr. (Southern California Edison Company, Glendale, Calif.): The paper under consideration discusses the organic-moderated reactor experiment (OMRE) conducted by the Atomics International Division of North American Aviation Inc., together with a description of their 12.5-emw (electrical megawatts) OMR, now being designed for the city of Piqua, Ohio, and their proposed 150-emw OMR plant for large central-station power plants.

It frankly presents both the advantages and problems associated with the organic fluids, which is the key factor in the successful and economical operation of the OMR concept.

The estimated cost figures presented in the paper, in the order of 18 and 10 mills per kilowatt-hour for small and large plants respectively, and which are based on a relatively short but highly encouraging operating experiment, appear to be conservatively reasonable and indicative of the justification for exploring the concept by constructing full-scale plants.

The results of the OMRE were somewhat unique in the history of power-reactor developments, in that the operating performance was in full accord with design specifications and was without unforeseen problems of significance. Undoubtedly, this was due in part to the compatibility of the organic with the materials of the fuel elements as well as with other conventional materials of construction.

The paper comments that an equilibrium polymer content up to 30% has been achieved and that the polymer formed at this concentration was $2^{1}/_{4}$ pounds per megawatt-hour. While total quantities of the organic are not given, it is presumed that $2^{1}/_{4}$ pounds per emw-hour make-up could result in a 100% replacement on an annual basis and is in support of achieving equilibrium over an extended period of time. Perhaps the authors would care to comment further on the expected performance of the organic over an extended period of time. This is very important since the estimated make-up now contributes 1 mill per electrical kilowatt-hour to the operating costs.

A second point the authors may wish to comment further on is the low steam pressure of 450 pounds per square inch. Since there is nothing inherent in the OMR design that prohibits the use of higher steam pressures, it would appear that in view of the steam temperature available, a higher pressure would produce a better heat rate and a higher thermal efficiency.

A third question of interest would be the expected consequences of a steam-generator failure whereby water might get into the primary loop.

Since the OMR for Piqua will utilize Santowax-R in place of Santowax-OM, and aluminum-clad fuel elements instead of stainless-steel cladding, a confirmation of the expected reduced rate of polymer formation, as predicted by reason of the reduced gamma exposure resulting from the lower fuel enrichment, will, no doubt, be a factor of interest in the operation at Piqua.

The load-change characteristics of the OMR operation at Piqua should also prove to be of interest. Although the controls will be designed to maintain constant steam pressure, they do appear to permit a temperature variation. Assuming the manufacturer's conservative loading schedule for steam turbines of 5% rated load per minute with constant temperature, it is apparent that with a temperature change of 75 degrees this load time could be increased appreciably.

In summary, the principal disadvantages of some instability of the organic under radiation coupled with poor heat-transfer properties with respect to water, this organic moderator and coolant enjoys impressive offsetting advantages in: 1. high-temperature capabilities at low pressure because of its low vapor pressure; 2. compatibility with fuel, canning, and structural materials resulting in low corrosion rates; 3. low induced radioactivity; and 4. no hazardous chemical reactions.

In addition to these characteristics either precluding the necessity for or substantially reducing the cost of external containment, it could conceivably result that this type of reactor plant would be acceptable closer to the load centers than would other types, with consequently reduced transmission cost and losses.

W. E. Parkins and E. F. Weisner: Since the paper was written, additional operating experience has been obtained on the OMRE. This reactor has operated for extended periods of time at 40% polymer content. There was a further decrease in the decomposition rate, and hence in the projected organic make-up costs, as expected. It now appears that for a power reactor of the Piqua type, in which slightly enriched uranium metal or oxide is used as fuel, that the organic make-up cost will be approximately 0.7 mill per electrical kilowatt-hour. This figure includes the effect of the absorption of gamma energy by the fuel itself, a point which the operation at Piqua will quantitatively establish.

The steam pressure of 450 pounds per square inch is used in the Piqua plant because all the installed boilers and turbines operate from a common steam header at this pressure. The nuclear reactor will supply steam to this header as do the other (conventional) boilers and the steam must therefore be produced at the same pressure. Organic-moderated reactors could be used to generate steam at much higher pressures if this is required.

The consequences of a steam-generator failure have been studied and devices have been provided to cope with such an event. A tube failure will cause a pressure increase on the organic side which will be relieved through two rupture-disk relief valves. The fluid discharged will be condensed and contained in the reactor drain tank. Operation of a pressure-relief valve will also provide a signal which shuts the reactor down and relieves the pressure on the steam side of the steam generator. Small water leaks in the steam generator can be tolerated since the degasifier provided to remove the radiolytic gases formed, will simultaneously remove any water present in the organic.

The reactor steam-generating complex will be operated to produce constant-pressure steam and also constant-temperature steam over a load range of 15% to 100% of full power. The rate of load change with this plant can be as rapid as with a conventional plant, being limited by the turbine in either case.

We wish to thank Mr. Werden for his comments on this paper and his appreciation of the problems of nuclear-reactor power plants.

◆

Discussion of "The Controlled Rectifier: Key to the Continuing Control Renaissance," by J. D. Harnden, Jr., published in Communication and Electronics, January 1959, pages 1006–12.

I. K. Dortort (I-T-E Circuit Breaker Company, Philadelphia, Pa.): In arriving at some estimate of the efficiency of equipments using the controlled rectifier *ZJ-39A*, the author points out that the forward drop is but one volt out of a peak voltage rating of 400. We assume the peak voltage rating is equivalent to the peak inverse voltage rating of power diodes, in which case it is expected that the recommended rms alternating voltage to be applied to these rectifiers would not be much over 115 volts. Even at this reduced voltage, the efficiency is still excellent. However, the author's statement raises the question of recommended peak operating voltage, a question which is further emphasized by the statement that it is possible to use these devices in a control system in which forward conduction is initiated by exceeding the "breakover" voltage instead of a signal on the control electrode. Since the forward breakover voltage and the peak inverse voltage rating of the rectifier are stated to be approximately equal, there would seem to be considerable danger of failure in applying controlled rectifiers in this manner. Perhaps the author would like to comment further on this point.

We are puzzled by the statement that time delay is necessary between the triggering of parallel connected controlled rectifiers. Mention is made of equalizing inductors as a suitable means for delaying the switching action. It would seem more reasonable to use these "equalizing inductors" as they are used with electronic devices to force simultaneous firing rather than purposely to delay some of the rectifiers and thereby cause them to carry less than their proportionate share of load.

The bridge-connected rectifier of Fig. 2 (H) is claimed to have inherent commutation protection. I do not understand what is meant here and would appreciate the author's explanation.

T. Bernstein (University of Wisconsin, Madison, Wis.): The literature dealing with controlled rectifiers emphasizes the fact that the gate circuit loses control after firing. The author mentions that there is a possibility that the controlled rectifier can be turned off by gate action. Have such units been made? What would be the magnitude of the gate signal necessary to turn off the rectifier as compared to the signal necessary to turn it on? Since the controlled rectifiers presently available are not designed for reverse voltage on the gate junction, I am in doubt as to whether the turn-off gate voltage would be a reversed polarity gate signal. This information will be useful as an indication of the future potentialities of the controlled rectifier.

◆

Discussion of "Magnet Wire for 600 C Temperature Transfers," by W. W. Wareham, published in Communication and Electronics, March 1959, pages 66–70.

Marvin M. Klein (Kollsman Instrument Corporation, Elmhurst, N. Y.): Mr. Wareham's paper is extremely interesting in view of some parallel problems which have been investigated here. We have been engaged in the development of rotating transducers for use at ultrahigh temperatures for the last year and a half. At that time the use of nickel-clad copper wire was considered and rejected for the following reasons:

1. An examination of the equilibrium diagram of nickel and copper indicates the existence of a continuous group of stable solid solutions from zero % of one of the elements to 100% of the element.

2. Inter-diffusion rates between atoms of the plate metal and that of the substrate metal are negligible at room temperatures. However, as is well known in kinetic theory, atomic or molecular motilities increase rapidly with temperature. Thus, at the 600 C temperatures referred to in the paper being discussed, the alloying process would take place in finite, appreciable time. Furthermore, the entropy of the process would indicate that the process would be complete only when the plate would be completely destroyed, and when a uniform composition would exist throughout the wire.

3. In general, alloying, whether due to substitutional solid solutions, interstitial solid solutions, or to intermetallic compounds, tends to increase the electrical resistance. This is understandable, if one considers the so-called "Free Electron" theory of solid conductors. In this theory the bond energies between the nuclei of the atoms making up the crystalline lattice structure, and some of the outer planetary electrons are low. Thus so-called free electrons are continuously migrating back and forth, as energy bonds are continuously being broken and reformed. It is the distribution of these free electrons which determines the comparative conductivity of the material. The introduction of any foreign atoms into a conductor's lattice structure usually distorts the lattice parameters enough to restrict the free electron motility. Thus, the resistance is permanently increased.

From the points brought out earlier in this discussion the resistance must continuously and irreversibly increase with time and temperature.

As a result of this line of thinking, we dropped the idea of using nickel-plated copper from further consideration, and decided to concentrate on finding some organization that would satisfactorily coat pure silver with ceramic or glass for us.

In further agreement with our thinking, a nearby manufacturer of fine wire reported the following to us some months ago:

He had coated fine silver wire with a light plating of aluminum and had then anodized the aluminum, thus securing a high conductive silver wire with a good ultrahigh temperature coating. Unfortunately, the diffusion affect at high temperatures and the consequent increase in resistance led to the dropping of this wire, at least in the fine sizes.

It might be added in conclusion that the "skin-effect" phenomenon makes this permanent increase of resistance particularly serious in the windings of transformers to be used for pulse transmission. In such applications rather serious degradation of leading edge might be expected.

Vol.
Pt. 1
195

Digital Computer Design of Pulse Transformers

P. E. LEGO
ASSOCIATE MEMBER AIEE

R. L. GREENE
ASSOCIATE MEMBER AIEE

J. M. BANIC, JR.
NONMEMBER AIEE

THE DEVELOPMENT of pulsed radar created a need for a relatively new type of transformer, one that transforms the energy in a pulse with an approximately rectangular shape and a time duration in the order of microseconds.[1] A good deal of experimental work has been done in this field, which has evolved a trial and error method of designing pulse transformers.[1-3] This method is tedious and time consuming since it involves the use of many interdependent mathematical calculations and the necessity of making these same calculations over and over again before a satisfactory design is obtained.

Different people have attempted to reduce the labor involved in the trial and error method through the use of nomograms and empirical charts prepared from past designs.[4] Although the nomogram approach to the trial and error method has helped reduce the number of trial calculations in the design procedure, pulse transformer designing still has required a disproportionate amount of calculations.

The availability of digital computers and their use in the solution of similar transformer design problems[5,6] suggested an investigation into the feasibility of utilizing digital computers to design pulse transformers.

The first problem was to choose a design method for programming the computer. There were at least three approaches to this problem: first, to use the customary trial and error method of the design engineer; second, to use the nomogram form of design procedure and reduce the nomograms to mathematical equations; and third, to try a completely new approach to the design problem. The decision was made to program the trial and error method of the design engineer. This method was chosen in order to utilize a proved technique with a minimum of programming and provide usable results in the immediate future. Incorporating the nomograms into the program would have reduced the number of iterations performed by the computer, but at the expense of more complex programming. The basic trial and error program would still have been required in either case, since the nomogram method is not complete in itself. A new design procedure would have delayed the program for an indefinite period of time, and was not economically feasible.

Since the specification information required for a pulse transformer is covered in Electronics Industries Association Standard *RS176*,[7] this was adopted as the major part of the input data to the computer.

In addition, it was felt that the input data should include information for setting the allowable tolerances on the electrical parameters required by the customer specifications.

Design Philosophy

The output of a typical pulse transformer resembles the waveform shown in Fig. 1. The objective of the transformer designer is to obtain a specified level of surge impedance while maintaining the rise time, fall time, and droop within guaranteed temperature rise. The trial and error method of design is the usual way of attaining these goals and tailoring the design for minimum size, weight, and economical manufacturing. This method involves assuming certain materials and configurations, and using recommended working stresses for safe practice under these conditions, calculating the performance parameters that can be realized. Comparing these parameters with those required to meet the customers' specifications reveals which, if any, of the initial assumptions must be modified to more nearly match these quantities. This process is continued until all the specifications are met, or until it can be demonstrated that the design requirements are contradictory and, therefore, impossible to attain.

In attempting to program the trial and error design method into the computer, a complete mathematical outline of the procedure was prepared along with a block diagram showing the necessary design steps. This block diagram is referred to as a flow diagram, since it shows the actual flow of information and decisions through the computer.

Initial Procedure

With the problem in an elementary flow diagram form, the next step was to adapt certain of the trial and error design procedures to computer usage.

1. The first step was to detail the input information necessary to produce a completed design. It was thought that the program should be self contained, or continuous, since intermediate observation of data is time consuming and costly.

2. The computer makes best use of information in mathematical equation form. Therefore all the data, whenever possible, were reduced to this form. Such data as standard parts had to be retained in tabular form. These data were reduced to a minimum and placed in the stored data area of the computer.

3. In order to have a foundation on which to base the program, certain assumptions have to be made such as the choice of working values for magnetic flux density, current density, and other factors which can be based on power or voltage levels.

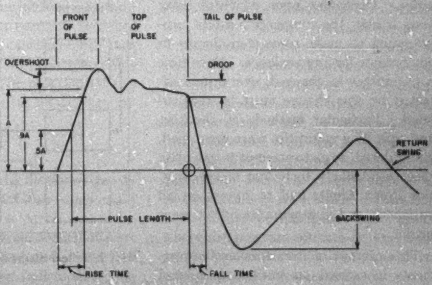


Fig. 1. Typical transformer output pulse. A is the nominal amplitude of the pulse

---

Paper 59-48, recommended by the AIEE Computing Devices and Electronics Committees and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted June 27, 1958; made available for printing November 21, 1958.

P. E. Lego, R. L. Greene, and J. M. Banic, Jr. are with the Westinghouse Electric Corporation; Mr. Lego is at East Pittsburgh, Pa., and Mr. Greene and Mr. Banic are at Greenville, Pa.

4. It is desirable to choose the proper winding scheme with which to begin the design. There are a number of such schemes that can be used by the computer in designing a pulse transformer. The winding schemes in the computer program are sufficient to handle most of the requirements desired by the customer. Each of these schemes has the lowest leakage inductance and capacitance product for a given volume of coil to insure good high-frequency response. The configurations differ in the range of surge impedances that can be realized, the preponderance of secondary current caused by bifilar heater power, the transformer turns ratios that are practical, and the polarity of the output voltage with respect to the input. The input data specifies the sequence of winding schemes which are used by the computer. If the surge impedance cannot be brought within specification using a particular scheme, the computer will automatically select a new scheme and begin a redesign. This process continues until the surge impedance is matched, or until all the winding schemes have been tried. The computer will automatically halt in the latter instance. However, the information which the computer has calculated to this point will be printed out. This essentially gives a family of designs which may be used at the designer's discretion.

5. The key parameter that generates a family of designs, clustered about the required performance, is the number of turns in the coil. The initial trial value is obtained from a turns-index formula. This value is altered on successive iterations, when necessary, to meet the performance specifications.

6. The design engineer usually possesses the ability to make a certain amount of intuitive corrections while designing a transformer. He can examine some particular quantity which he had calculated, and through past experience, realize that this value is out of line. Or, he can project his thinking to future calculations and foresee that the design is going to fail in some respect. It is extremely difficult to program these engineering abilities into the computer. The machine can make only logical, yes or no, type decisions. It is therefore necessary to categorize every check that is made, on any calculation, into a logical type operation.

A quantity is either greater than, equal to, or less than a specified guarantee. Depending upon which of these tests is met, the computer must be programmed to make some step change to compensate for any variance. Each time a parameter is changed, every item affected by this change must be recalculated. Particular methods of checking and matching quantities were developed, and ordered, so as to attempt to calculate those most likely to be out of specification first. Limits had to be set on all variances and compensations that were made.

This process of iterations and adjustments is carried on until a completed design is obtained. The information obtained on the intermediate designs is also retained by printing it on magnetic tape before beginning a new iteration.

The initial outline and flow diagram were continually tested by hand calculating sample transformer designs. Corrections and additions were made where necessary. Out of this procedure evolved the completed flow diagram shown on Fig. 2.

Computer Program

Having drafted the final flow diagram, the next step was to program a digital computer to perform the necessary calculations. The IBM (International Business Machine Corporation) *704* digital computer was chosen for this project because of its availability and also because a large internally stored computer was desired.

This computer has three modes of information storage. These are: magnetic-core storage, magnetic-drum storage, and magnetic-tape storage. The magnetic-core storage is the primary storage unit of the computer and can retain information indefinitely. Information can be recalled from this type of storage in a few millionths of a second. Only information stored in this area can be directly used for calculation by the computer. The magnetic tapes supply the means for storing large quantities of information until it is needed. It must then be brought into the magnetic-core storage area for direct usage. Of these three modes, only magnetic-core and magnetic-tape storage were used in programming the pulse transformer design.

PROGRAM FORMULATION

The program was written completely in the computer's own language and introduced into the calculator by key punched instructions on standard IBM cards. The magnetic-core storage was not large enough to hold all the data and instructions necessary to cause an uninterrupted solution to this problem. For this reason, the program was divided into five phases. These phases are

1. An input routine which has the major function of loading data peculiar to each particular transformer solution into the calculator. In addition, this portion of the program contains all of the data used in common by the other sections. This data is comprised of such items as numerical constants, SHARE subroutines for finding square roots, natural logarithms, and exponentials, and other information such as wire tables and insulation thicknesses.

2. The major calculating routine contains all of the calculations normally performed in obtaining a transformer design, except those needed for finding transformer response.

3. The performance calculating routine, which finds such factors as rise time, fall time, droop, and overshoot by solving the differential equations generated by the transformer equivalent circuit. All of the operations needed to obtain complete performance data for the problem are performed here.

4. The output routine, which prints out all of the information calculated while the computer is iterating to a finished design. This routine prints out the response data needed to obtain a plot of the output pulse. It also prints out all of the transformer parameters as well as information on losses, efficiency, copper weights, and other pertinent data.

5. A master routine which controls all of the other routines and dictates the complete design process. It is this routine which contains the instructions needed to bring in the other routines from tape at the correct time. Each of the preceding phases was developed separately. The calculating routine and the transformer response sections were completed first. The input and output routines were written next; and the master routine last. Except for the master routine, all of

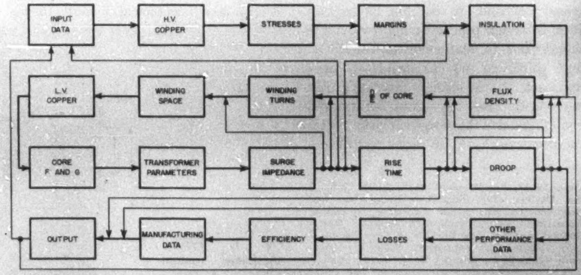


Fig. 2. Flow diagram. D and E are the core area proportions. F and G are the core window proportions

the preceding phases are stored on a single magnetic tape called the program tape. The master routine is stored on ten punched IBM cards. The specification input data needed for each transformer design is also punched onto IBM cards.

To execute a design, the master routine deck is read into the calculator by means of a card reader. The input data cards are placed in the card-reader hopper. The master routine then takes complete control of the proceedings. It first brings the input routine into magnetic-core storage from the program tape. The input routine causes the data cards to be read from the card hopper and placed in core storage. It also brings the stored data necessary for the design into core storage and returns the program control to the master routine. The master routine then manipulates the program tape in such a manner as to have the proper routine in core storage at the proper time. This process is continued until a finished design is reached. When a design is completed, the master routine will once again bring the input routine into core storage. The input routine will sense the card reader and attempt to read a new set of data cards. If there are more data cards present, another design will be initiated. If there are no more data cards in the hopper, then the computer will automatically halt.

Debugging

The debugging, or removing of errors from the program, was performed in stages. Each of the individual routines was tested separately using the data obtained while hand testing the flow diagrams. This procedure was used to ensure that the mathematical outline was being followed correctly. Also, it served as a check on the flow pattern of the routines. Once the individual routines had been successfully tested, the master program was written. This was the link needed to tie the individual sections together and enabled testing to start on the over-all program. Computer design of pulse transformers, which had already been hand designed, were attempted next. This procedure checked the program for logical correctness and allowed program changes to be made in order to improve the transformer design. During the debugging phase, the individual routines were stored on punched cards to facilitate the correction of errors. The entire program was placed on magnetic tape after the debugging process was completed. The master routine also handles this function. At the present, any new program changes required are inserted into the punched card deck and a new program tape is written.

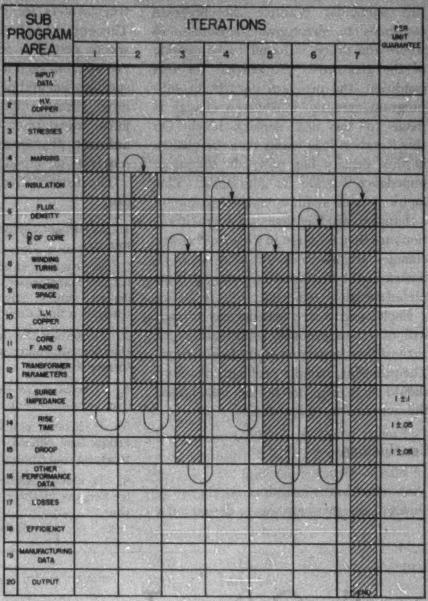


Fig. 3. Sample problem

The Completed Program

The final form of the program contains a total of over 7,000 machine instructions and 500 pieces of stored data. All of this information is stored on a single magnetic tape, except for the master program which is punched on ten standard IBM cards. The program is completely self contained in that once the design process is started, it continues without intervention until a completed transformer design is obtained.

To run a design it is only necessary to read the master routine into the calculator and place input data cards in the card reader hopper. The calculator will then complete the design and print the results on magnetic tape. The program takes a few minutes to complete an average design.

Sample Design

The sample problem displayed graphically in Fig. 3 shows the flow which might occur in designing a particular transformer. The subprogram areas shown are the same as those displayed in the flow diagram. It might be well to state that these areas are not completely inclusive and it is possible that iterations may be required within a subprogram to produce the desired results. The flow for this problem proceeds directly to area no. 13. Here, the computer checks the calculated surge impedance against the guarantee. The calculated value is found to be outside of the guarantee range and some type of compensation is required to correct for this variance. From the flow diagram of Fig. 2, we see that there are several paths which the computer can take from this point. The path chosen depends on whether the surge impedance is below, or above, the guarantee and also upon how great the deviation is. The path shown in the sample problem is that which would be taken if the surge impedance were lower than the guarantee but within a specified percentage of the guarantee. The program here retraces to subprogram area no. 5 where the insulation thickness would be adjusted by a ratio calculated from the magnitudes of the calculated surge impedance and its guarantee. Here, the program iterates to subprogram area no. 8 where the winding turns would be changed. Once again, where the program retraces to, and what particular compensation is used, is determined by tests which are made on the calculated values and the guarantee. For this particular example the program continues to subprogram area no. 15 where the droop is found to be outside of the guarantee. To correct for this condition, the program iterates to subprogram area no. 6 where a change is made in the flux density level. On this iteration it is found that the change in flux density has now put the surge impedance outside its guarantee. This factor is once again compensated for by a change in winding turns. Several iterations are then required in which successive changes are made in the ratio of D to E and in the flux density, before a completed design is achieved.

Each particular transformer design will, of course, probably follow a different path of flow in getting to its final design. The preceding sample problem, however, should help one to better understand the complexities of any design.

Conclusions

As stated in the introduction, the initial object of this program was to design pulse transformers using a digital computer. This object has been met and in most respects surpasses what was initially expected from such a program. With only specification and performance data for an input, the computer completely designs a pulse transformer. This design is done automatically in a continuous process, requiring no outside monitoring.

The distinct advantages of the program are:

1. It reduces the design time from days to minutes.
2. It provides, in addition to a final design for the transformer, all of the intermediate design values calculated while iterating towards a final design.
3. It enables the design engineer to study the effects of parameter variation on transformer design.
4. It releases the design engineer from the drudgery of routine calculations.
5. It insures consistency of design procedure and should allow the manufacturer to give rapid quotes on proposed designs.
6. It should allow faster delivery service because of the reduction in engineering time.

References

1. Pulse Generators—Radiation Laboratory Series, G. N. Glasoe, J. W. Lebacqz. McGraw-Hill Book Company, Inc., New York, N. Y., vol. V, 1948.

2. High Power-High Voltage Pulse Transformers—Design Criteria and Data, P. Fenoglio, C. W. Peck, F. R. Richardson, H. W. Lord, A. Boyajian. *Final Report*, General Electric Company, Pittsfield, Mass., 1953.

3. Pulse Transformer Design and Test Methods, P. R. Gillette, K. W. Henderson, K. Oshima, R. M. Rowe. *Final Report*, Project 782, Contract no. *DA-36-039 SC 42645*, Stanford Research Institute, Stanford, Calif., "Part I—Test Procedures for Pulse and Broadband Transformers"; "Part II—Design Procedures for Pulse Transformers"; "Part III—Description of Research," 1955.

4. Pulse Transformer Design Chart, Reuben Lee. *Electronic Equipment*, New York, N. Y., 1957.

5. Application of a Digital Computer to the Design of Power Transformers to Specification, C. L. Moore, W. T. Duboc, P. A. Zaphyr. *AIEE Transactions*, vol. 75, pt. I, May 1956, pp. 134–38.

6. Computer Design of Small Electronic Transformers, W. Etchison, M. B. Meunier, R. Lee. *Transactions*, Professional Group on Component Parts, Institute of Radio Engineers, New York, N. Y., 1958.

7. Pulse Transformers for Radar Equipment. *ETA Standard RS176*. Electronic Industries Association, New York, N. Y.

The Optimum Transmission-Line Pulse Transformer

F. J. YOUNG
ASSOCIATE MEMBER AIEE

E. R. SCHATZ
MEMBER AIEE

J. B. WOODFORD
ASSOCIATE MEMBER AIEE

TAPERED transmission lines have been found to be useful for pulse-transformer service.[1,2] It is the object of this paper to develop a new theory for the optimization of the taper which is based upon the transient rather than the steady-state response of the transmission-line pulse transformer. A criterion for the optimum taper will be established and then applied in a special case in which both ends of the line are nominally matched. Finally, the case of general terminations at both the input and output will be discussed.

There have been numerous analyses of the optimum taper based on the steady-state response of the tapered transmission-line pulse transformer. For instance, Bolinder establishes a Fourier transform method for the approximate calculation of a quantity defined as a reflection coefficient which has the same mathematical form as the reflection coefficient commonly defined for a uniform line. Optimum tapers are sometimes found by minimizing this reflection coefficient.[3] Lewis and Wells present expressions for both Bolinder's and a conventional reflection coefficient which are the same under certain conditions.[4] Bolinder's theory is not very useful for transient analyses because it does not fully consider the influence of the terminations and is based upon steady-state considerations. However, when the internal impedance of the sending-end signal generator equals the complex conjugate of the nominal sending-end impedance, Bolinder's reflection coefficient (evaluated at the sending end) takes on a physical meaning. It is a direct measure of how well the signal generator is matched to the sending end of the loaded line.

Many later works published in this field have been devoted to finding tapers which would minimize Bolinder's reflection coefficient over various ranges of frequency.[5–8] Physically, minimization of this coefficient results in maximum power being accepted by the input end of the tapered transmission line. For steady-state use, this is sufficient to guarantee maximum power at the load because all power which is put into the line eventually arrives at the load if the line is without loss. However, for pulse work it is necessary that the input power arrives at the load when it is needed, rather than at some later time. If some portion of the spectrum of input power is allowed to pass through the line more quickly than another portion, severe distortion may result in load voltage. Fidelity of pulse response is, of course, a criterion for optimization. Because the minimization of Bolinder's reflection coefficient does not necessarily insure a good reproduction of the input waveform at the output of a line, such a procedure does not yield the best taper

---

Paper 59-16, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted May 14, 1958; made available for printing November 12, 1958.

F. J. Young, E. R. Schatz, and J. B. Woodford are with the Carnegie Institute of Technology, Pittsburgh, Pa.

This work was supported by United States Navy Contract No. *760(09)*.

Vol.
Pt. 1
1959

for a transmission-line pulse transformer.

The new theory developed here requires an accurate, usable transient solution for tapered transmission lines. Such a solution for all lines having a constant velocity of propagation has been derived in a companion paper.[9]

Initial Slope Criterion

The initial slope criterion was used by Kukel for judging the performance of various types of exponential-line pulse transformers. This criterion is a direct measure of how much distortion a short pulse will suffer when transformed in amplitude by a tapered line.[10] For example, let the transient responses sketched in Fig. 1 be the results of applying equal-step voltages to two equal-length lines having the same velocity of propagation, ratio of impedance transformation, nominal input and output impedances and terminations, but obeying different taper laws. The dotted curves are the responses of the lines to a pulse of length τ. It is apparent that line II distorts pulses more than line I. This is so because the magnitude of the initial slope

$$\left.\frac{dV}{dt}\right|_{t=\frac{L}{c}}$$

of the step response of line II is greater than that of line I. Thus the initial slope criterion can be stated as follows. In the case of lines having the same electrical length, the same ratio of impedance transformation, the same input and output nominal impedances, the same terminations, the best line for transforming pulses, which are short compared to the time it takes the step function response of that line to approach its final value, is that line which has the smallest magnitude of intial slope of the response to an exciting step function.

Derivation of the Optimum Line When Both Ends of the Line Are Nominally Matched

The initial slope μ of the output of a tapered line in finite difference form is

$$\left.\frac{V_{t+\Delta t}-V_t}{\Delta t}\right|_{t=L/c}$$

The quantity Δt can be chosen for convenience. Let $\Delta t=2L/cn$ where n is the number of elementary steps in the line. By equation 8 of the previous paper[9] $V(L/c)=K$ and $V(L/c+2L/cn)=K(1-m_1)$.

Then

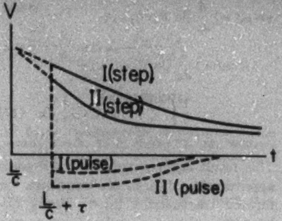


Fig. 1. Typical responses of a tapered transmission line to a step and a pulse

$$|\mu|=\left|\frac{cnm_1K}{2L}\right| \tag{1}$$

In the Appendix it can be seen that

$$m_1=\sum_{j=1}^{n-2}\rho_j\rho_{j+1}$$

for lines which are nominally matched at both ends. It is also shown in the Appendix that for large values of n

$$\sum_{j=1}^{n-2}\rho_j\rho_{j+1}=\frac{1}{4n}\int_0^L\left\{\frac{d}{dx}\,[\ln R(x)]\right\}^2dx \tag{2}$$

Substitution of equation 2 into equation 1 yields

$$|\mu|=\left|\frac{cK}{8L}\int_0^L\left\{\frac{d}{dx}\,[\ln R(x)]\right\}^2dx\right| \tag{3}$$

Because c, K, and L are constants the only quantity which can be adjusted to minimize $|\mu|$ is the taper function $R(x)$. The function $R(x)$ which makes the integral I a minimum must be determined. This will be done by the methods of the Calculus of Variations.[11,12]

Let

$$2I=\int_0^L\left\{\frac{d}{dx}\,[\ln R(x)]\right\}^2dx$$

and $R(x)=y+\alpha n(x)$ where $n(x)$ is any function such that $n(0)=n(L)=0$. Then

$$2I(\alpha)=\int_0^L\left\{\frac{d}{dx}[\ln(y+\alpha n[x])]\right\}^2dx \tag{4}$$

Differentiating $I(\alpha)$ with respect to α yields

$$I'(\alpha)=\int_0^L\left\{\frac{d}{dx}[\ln(y+\alpha n[x])]\right\}\times\left\{\frac{d}{dx}\left[\frac{n[x]}{y+\alpha n[x]}\right]\right\}dx \tag{5}$$

It can be seen that

$$I'(0)=\int_0^L\left\{\frac{d}{dx}\,[\ln y]\right\}\left\{\frac{d}{dx}\left[\frac{n(x)}{y}\right]\right\}dx \tag{6}$$

This expression can be integrated by parts. Let $U=d/dx(\ln\ y)$ and $dV=d/dx[n(x)/y]$.

From integration by parts

$$I'(0)=\left.\frac{n(x)}{y}\,\frac{d}{dx}(\ln y)\right|_0^L-\int_0^L\left[\frac{n(x)}{y}\,\frac{d^2}{dx^2}(\ln y)\right]dx \tag{7}$$

But by definition $n(0)=n(L)=0$. Under this condition equation 7 becomes

$$I'(0)=-\int_0^L\left[\frac{n(x)}{y}\,\frac{d^2}{dx^2}\,(\ln y)\right]dx \tag{8}$$

In order to find the function y which makes $I(0)$ an extremum, $I'(0)$ must be set equal to zero. For this to be true the integrand of equation 8 must equal zero. Because of the nature of the functions involved, the integrand is zero only if

$$\frac{d^2}{dx^2}(\ln y)=0 \tag{9}$$

The solution to this equation is given by $y=\exp(cx+b)$. But because $\alpha=0$, $R(x)=y$. To find the constants c and b the boundary conditions at $x=0$ and $x=L$ must be used. Upon the evaluation of c and b, $R(x)$ becomes

$$\frac{R(x)}{R(0)}=\exp\left\{\frac{x}{L}\left[\ln\frac{R(L)}{R(0)}\right]\right\} \tag{10}$$

Thus, for the case of a line nominally matched at both ends, the optimum taper is exponential. This is not too surprising in view of the results obtained in the previous paper.

The Optimum Line in General

The most general case that can be considered by the theory developed in this paper is that of finding the optimum tapered line in terms of its terminations. For this case an expression for the voltage response of a line with arbitarary terminations must be derived. Equation 20 of the appendix gives the most general form of m_1. The quantities ρ_n and ρ_0 are functions of the terminations and ρ_j is defined in the Appendix. When n gets large the quantities ρ_j become proportional to the logarithmic derivative of the surge resistance function $R(x)$. For large values of n, m_1 is given by

$$m_1=\frac{\rho_nd}{2ndx}[\ln\ R(x)]_{x=0}+\frac{\rho_0d}{2ndx}[\ln R(x)]_{x=L}+\frac{1}{4n}\int_0^L\left\{\frac{d}{dx}[\ln\ R(x)]\right\}^2dx \tag{11}$$

Substitution of m_1 into equation 2 of the previous paper yields the response of the line after much algebraic manipulation.

$$V(s,L)=\frac{\frac{1}{2}V_g(s)\epsilon^{-sL/c}\frac{2R_n}{R_n+Z_g}\frac{2Z_L}{R_1+Z_L}\prod_{j=1}^{n-1}\frac{2R_j}{R_j+R_{j+1}}}{1+\frac{1}{2n}\left\{\rho_n(s)\frac{d}{dx}[\ln R(x)]_{x=0}+\rho_0(s)\frac{d}{dx}[\ln R(x)]_{x=L}+\frac{1}{2}\int_0^L\left[\frac{d}{dx}\ln R(x)\right]^2dx\right\}\epsilon^{-2sL/nc}+\ldots} \tag{12}$$

Let

$$V_g(s)=\frac{e_g}{s}\text{ and }K=2e_g\prod_{j=1}^{n-1}\frac{2R_j}{R_j+R_{j+1}}$$

Then by division

$$\frac{V(s,L)}{K}=\frac{R_nZ_L(s)\epsilon^{-sL/c}}{s[R_n+Z_g(s)][R_1+Z_L(s)]}\times\left(1-\frac{\epsilon^{-2sL/cn}}{2n}\left\{\rho_n(s)\frac{d}{dx}[\ln R(x)]_{x=0}+\rho_0(s)\frac{d}{dx}[\ln R(x)]_{x=L}+\frac{1}{2}\int_0^L\left[\frac{d}{dx}\ln R(x)\right]^2dx\right\}+\right) \tag{13}$$

Equation 13 is the Laplace transform of the output of a tapered line with taper function $R(x)$ terminated at the input with an impedance $Z_g(s)$ and at the output with $Z_L(s)$. By the initial value theorem of Laplace transform theory

$$V\left(\frac{L}{c},L\right)=\frac{KR_nZ_L(\infty)}{[R_n+Z_g(\infty)][R_1+Z_L(\infty)]} \tag{14}$$

$$V\left\{t=\frac{L}{c}\left[1+\frac{2}{n}\right],L\right\}\cong\mathcal{L}^{-1}\left\{\frac{KR_nZ_L(s)\epsilon^{-sL/c}}{s[R_n+Z_g(s)][R_1+Z_L(s)]}\right\}\Bigg|_{t=L/c(1+2/n)}-\frac{KR_nZ_L(\infty)\left\{\rho_n(\infty)\frac{d}{dx}[\ln R(x)]_{x=0}+\rho_0(\infty)\frac{d}{dx}[\ln R(x)]_{x=L}+\frac{1}{2}\int_0^L\left[\frac{d}{dx}\ln R(x)\right]^2dx\right\}+\ldots}{2n[R_n+Z_g(\infty)][R_1+ZL(\infty)]} \tag{15}$$

If n is large the first term of equation 15 will be almost equal to the quantity $V(L/c,L)$ given by equation 14. By substituting equations 14 and 15 into equation 1 the initial slope can be determined. It is given by

$$\mu=\frac{cnm_1V\left(\frac{L}{c},L\right)}{2L} \tag{16}$$

Given the terminations, the length of line, and the nominal impedances at each end, the taper function $R(x)$ which is optimum, can be found by determining the $R(x)$ which minimizes the magnitude of the initial slope μ. To minimize the initial slope, the function ϕ must be minimized. The function ϕ is given by

$$\phi=\rho_n(\infty)\frac{d}{dx}[\ln R(x)]_{x=0}+\rho_0(\infty)\frac{d}{dx}[\ln R(x)]_{x=L}+\frac{1}{2}\int_0^L\left\{\frac{a}{dx}[\ln R(x)]\right\}^2dx \tag{17}$$

This result is given support by Frank's approximate theory of transients on tapered transmission lines. Frank obtains the distortion suffered by a step function which has been transformed by a tapered line which is nominally matched at the load end.[13] The problem of minimizing Frank's distortion is identical to that of minimizing ϕ when $\rho_0(\infty)=0$. Frank's theory is valid for values of time very close to $t=L/c$ only.

The authors have not succeeded in obtaining an analytical solution for the $R(x)$ function in terms of the terminations which yields a minimum value of ϕ. Actually for some capacitative terminations the initial slope may have no minimum value and may take on positive values. However, equation 17 is of some computational value. For given terminations, various known tapers can be substituted into equation 17 to obtain the line with the most desirable initial slope.

Equation 16 is of value in another way. For appropriately short pulses the initial value of the output voltage from equation 14 and the initial slope from equation 16 may give enough of the response curve to analyze a line with given terminations and taper function.

Conclusions

The initial slope criterion is useful only during the interval of time that the step response falls off linearly with its initial slope. After this interval of time the initial slope criterion fails to yield any information about the transient response. By the initial slope criterion the lines considered in the previous paper in Fig. 5 are ordered from best to worst as follows: parabolic, hyperbolic, exponential, and Bessel. This choice remains valid until about $ct/L=1.1$. For pulses longer than that but shorter than $1.6L/c$ they are ordered from best to worst as follows: hyperbolic, exponential, parabolic, and Bessel. For still longer pulse lengths the order is different. Thus, it can be seen that the initial slope criterion must be used only in conjunction with the total transient response and with certain pulse lengths in mind. Clearly, if the initial slope criterion is used alone erroneous judgement of lines can occur. For example, if one wished to transform pulses of duration of 1.5 L/c seconds with one of the lines mentioned in Fig. 5 of the previous paper, blind application of the initial slope criterion would dictate the use of the parabolic line. As mentioned previously this would be a poor choice of line taper.

The criterion which is always valid is more like an average slope criterion which because of the complexity of the transient response of tapered lines can be applied more easily by visual or graphical means than analytically.

Appendix

The quantity n_1 is defined by equation 6 of the companion paper. In Appendix III of that paper it was found that

$$\alpha_1=\rho_{n-1}+\sum_{j=0}^{n-2}\rho_j\rho_{j+1}\text{ and}$$

$$b_1=\rho_{n-1}-\sum_{j=0}^{n-2}\rho_j\rho_{j+1}$$

For that reason the value of n_1 is identical for both the approximate and the exact transient solution to the step line. Hence, by equation 2 of the companion paper

$$n_1=\rho_0\rho_1+\rho_n\rho_{n-1}+\sum_{j=1}^{n-2}\rho_j\rho_{j+1}$$

By division

$$m_1=\rho_0\rho_1+\rho_n\rho_{n-1}+\sum_{j=1}^{n-2}\rho_j\rho_{j+1} \tag{18}$$

If the number of steps used to approximate the tapered line is large enough $\rho_j\cong\rho_{j+1}$. Then equation 18 becomes

$$m_1=\rho_0\rho_1+\rho_n\rho_{n-1}+\sum_{j=1}^{n-2}\rho_j^2$$

where

$$\rho_j=\frac{R_j-R_{j+1}}{R_j+R_{j+1}}\text{ for }0<j<n$$

The coefficient ρ_j may be written

$$\rho(x) = -\frac{R(x+\Delta x)-R(x)}{R(x+\Delta x)+R(x)}$$

where x is a measure of position along the line. Clearly

$$\underset{\Delta x \to 0}{\text{Limit}}\ \rho(x) = \frac{-1}{2}\frac{d}{dx}[\ln R(x)]$$

Since $\Delta x = L/n$, $n \to \infty$ when $\Delta x \to 0$. Under such a condition it can be shown that

$$\sum_{j=1}^{n-2} \rho_j^2 = \frac{1}{4n}\int_0^L \left\{\frac{d}{dx}[\ln R(x)]\right\}^2 dx \qquad (19)$$

Therefore

$$m_1 = \rho_0\rho_1 + \rho_n\rho_{n-1} + \frac{1}{4n}\int_0^L \left\{\frac{d}{dx}[\ln R(x)]\right\}^2 dx \qquad (20)$$

where

$$\rho_0 = \frac{Z_L - R_1}{Z_L + R_1} \text{ and } \rho_n = \frac{R_n - Z_g}{R_n + Z_g}$$

References

1. Pulse Transients in Exponential Transmission Lines, E. R. Schatz, E. M. Williams. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 38, 1950, pp. 1208–12.

2. Design of Exponential Line Pulse Transformers, E. M. Williams, E. R. Schatz. *Ibid.*, vol. 39, 1951, pp. 84–6.

3. Fourier Transforms in the Theory of Inhomogeneous Transmission Lines, F. Bolinder. *Transactions*, Royal Institute of Technology, Stockholm, Sweden, no. 48, 1951.

4. Millimicrosecond Pulse Techniques (book), I. A. D. Lewis, F. H. Wells. Pergamon Press Inc., New York, N. Y., vol. 5, 1954, pp. 63–93.

5. The Hyperbolic Transmission Line as a Matching Section, H. J. Scott. *Proceedings*, Institute of Radio Engineers, vol. 41, 1953, pp. 1654–57.

6. Parabolic Transmission Line, R. Yaug. *Ibid.*, vol. 43, 1955, p. 1010.

7. A Transmission Line Taper of Improved Design, R. W. Klopfenstein. *Ibid.*, vol. 44, 1956, pp. 31–5.

8. The Optimum Tapered Transmission Line Matching Section, R. E. Collin. *Ibid.*, vol. 44, 1956, pp. 539–57.

9. The Transient Response of Tapered Transmission Lines, F. J. Young, E. R. Schatz, J. B. Woodford. *AIEE Transactions*, see pp. 223–28 of this issue.

10. Engineering Problems in the Exponential Transmission Line Pulse Transformer Design, J. Kukel. *Doctoral Dissertation*, Carnegie Institute of Technology, Pittsburgh, Pa., 1953.

11. Differential and Integral Calculus (book), R. Courant. Interscience Publishers Inc., New York, N. Y., vol. II, 1950, p. 491.

12. Introduction to the Calculus of Variations (book), W. E. Byerly. Harvard University Press, Cambridge, Mass., 1917.

13. Reflections from Sections of Tapered Transmission Lines and Waveguides, N. Frank. *Report, no. 17*, Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., 1943, set. 43.

The Transient Response of Tapered Transmission Lines

F. J. YOUNG
ASSOCIATE MEMBER AIEE

E. R. SCHATZ
MEMBER AIEE

J. B. WOODFORD
ASSOCIATE MEMBER AIEE

TAPERED transmission lines are frequently used in communications engineering to transform voltages and impedances. Much has been done in the steady-state analysis[1–12] of these devices. The need remains, however, for a theory of the optimum taper based upon transient considerations because of intrinsic limitations of any steady-state theory. Steady-state treatments are sufficient for a-c applications, but are difficult to apply meaningfully to transient phenomena. Very little has been done in the realm of tapered-transmission line transient analysis. An analysis for exponential tapers has appeared in the literature,[13–16] but there has been no attempt to find the optimum taper from transient considerations.

A criterion for the judgement of various tapered lines can be established on the basis of their performances as pulse transformers. This criterion is easily set up by considering the distortion which a pulse suffers when it is transformed by a tapered line. The main problem of this paper is the determination of the transient response of the tapered line as a function of its terminations and nominal characteristic impedance. A companion paper considers the optimization problem.[17]

In this paper a general expression for the transient response of any tapered transmission line with a constant velocity of propagation to a step-function excitation is derived. To obtain such an expression an aggregation of ordinary transmission lines connected end to end is considered. An exact expression for the response of that circuit is obtained along with an approximate solution which is easier to apply and quite accurate under certain conditions. Both expressions can be written in closed form but their numerical evaluation is long and tedious. For that reason the IBM (International Business Machine Corporation) *650* computer is used and special care is taken to put the analytical solution into a form amenable to digital computation. The transient solution is evaluated for Bessel, hyperbolic, parabolic, and exponential (see Appendix I) lines for two cases: 1. for nominally matched lines at both the input and output, and 2. for lines excited by a generator of zero internal impedance with the load nominally matched.

The Step Line Method of Calculating the Transient Response

For simplicity the tapered transmission lines considered herein are approximated by a large number of cascaded uniform lines of differing characteristic impedances. These lines are joined together in such a manner that the characteristic impedance seen at any position along the assemblage of lines (which will hereafter be called a step line) approximates the nominal characteristic impedance at the corresponding position on the tapered line. Fig. 1 shows how the characteristic impedance of a tapered line can be approximated by the characteristic impedance of a step line. Let the sections of uniform line nearest the sending and receiving ends of the step lines be denoted by n and 1 respectively. The method of finding the transient response of such a step line is outlined in the following; see Fig. 6.

A. 1. Find the impedance Z_{n-1} seen looking into the $(n-1)$th section toward the 1st section.

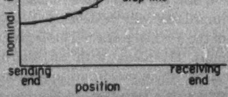


Fig. 1. Characteristic impedance of a step and tapered line versus position along line

---

Paper 59-15, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted May 14, 1958; made available for printing November 12, 1958.

F. J. Young, E. R. Schatz, and J. B. Woodford are with Carnegie Institute of Technology, Pittsburgh, Pa.

2. Use this impedance Z_{n-1} as a load for the nth section.
3. Apply a signal V_n at the sending end and determine by ordinary transmission line theory the voltage V_{n-1} which appears across the load Z_{n-1} of the nth section.

B. 1. Decrease the subscripts on all voltages and impedances and repeat the operations described previously until the voltage at the receiving end V_0 is found.

Fortunately this method can be applied to a step line of any number of steps or sections and it can be shown that as the number of steps is increased without bound the step line and the tapered line become identical.[18]

In Appendix II a voltage $V_g\ (s,0)$ is supplied to an n-step line by a signal generator whose internal impedance is $Z_g\ (s)$. The step line is terminated in an impedance $Z_L\ (s)$. Using the method outlined in the foregoing the output voltage $V\ (s,L)$ which appears across $Z_L(s)$ is calculated and after much algebraic simplification it becomes

$$V(s, L) = \frac{2V_g(s,0)\epsilon^{\frac{-sL}{c}} \dfrac{R_nR_{n-1}}{R_n+Z_g(s)} \displaystyle\prod_{j=0}^{n-2} \frac{2R_j}{R_j+R_{j+1}}}{\alpha_{n-1}(s)R_{n-1}\left[1+\rho_n(s)e^{-\frac{2sL}{cn}}\right]+ \beta_{n-1}(s)R_n\left[1-\rho_n(s)\epsilon^{\frac{-2sL}{cn}}\right]} \tag{1}$$

where R_j = the surge resistance of the jth section of line for $1 \leq j \leq n$.

$R_0 = Z_L(s)$

$$\rho_n(s) = \frac{R_n - Z_g(s)}{R_n + Z_g(s)}$$

s = the complex frequency
L = total length of the line
n = the number of sections in the line
$R_{n-1}\alpha_{n-1}(s)$ = the numerator of the impedance seen looking into the $n-1$ section of line toward $Z_L(s)$; see equation 20 of Appendix III)
$\beta_{n-1}(s)$ = the denominator of the same impedance
c = the velocity of light

The inverse Laplace transformation of equation 1 yields the exact transient response of the n-step line under consideration. Because equation 1 contains the numerator and the denominator of the impedance seen looking toward the receiving end from the $n-1$th section, a method for the exact determination of this impedance will be developed for the case of the $n-1$ step line. This is given in Appendix III along with an alternate derivation which yields an approximate impedance of a step line.

Equation 1 is the exact transient response of an n-step line. An approximate transient response for such a line can be obtained by substituting equation 23 of Appendix III into equation 1. The result is

$$V(s, L) = \frac{1}{2\xi} V_g(s, 0)\epsilon^{\frac{-sL}{c}} \prod_{j=0}^{n} \frac{2R_j}{R_j+R_{j+1}} \tag{2}$$

where

$$\xi = \left[1+\rho_n\rho_{n-1}\epsilon^{-\frac{2sL}{cn}}\right] \times \left[1+\sum_{i=1}^{n-1}\sum_{j=0}^{n-i-2}\rho_j\rho_{j+i}\epsilon^{-\frac{2Lis}{cn}}\right] + \left[\rho_{n-1}+\rho_n\epsilon^{-\frac{2sL}{nc}}\right]\sum_{j=1}^{n-1}\rho_{n-i-1}\epsilon^{-\frac{2siL}{cn}}$$

If $\rho_n = 0$, equation 2 becomes

$$V(s, L) = \frac{\frac{1}{2} V_g(s, 0)\epsilon^{\frac{-sL}{c}} \displaystyle\prod_{j=0}^{n} \frac{2R_j}{R_j+R_{j+1}}}{1+\displaystyle\sum_{i=1}^{n-1}\sum_{j=0}^{n-i-1}\rho_j\rho_{i+j}\epsilon^{\frac{-2Lis}{cn}}} \tag{3}$$

Equation 3 applies when the signal generator impedance equals the nominal sending-end impedance of the line.

Calculated Transients on Various Lines

It is convenient to calculate the step-function responses of various 100-step lines because such lines approximate tapered lines very well, and the resulting equations lend themselves to quick calculation on an IBM *650* digital computer. Because the responses of these lines are to be compared critically, it is necessary in order to reach sound conclusions that the lines should: 1. be of identical length, 2. possess the same ratio of nominal impedance transformation, 3. be terminated identically, and 4. have applied to them the same input. All lines herein considered have identical nominal input and output surge resistances and are terminated at the output in the nominal output surge resistance. Under the foregoing conditions and provided the internal impedance of the input signal generator equals the input and surge resistance, equation 1 becomes

$$V(s, L) = \frac{V_g(s, 0)\epsilon^{\frac{-sL}{c}} R_{n-1} \displaystyle\prod_{j=1}^{n-2} \frac{2R_j}{R_j+R_{j+1}}}{\alpha_{n-1}(s)R_{n-1}+\beta_{n-1}(s)R_n} \tag{4}$$

The approximate solution given by equation 2 becomes

$$V(s, L) = \frac{\frac{1}{2} V_g(s, 0)\epsilon^{\frac{-sL}{c}} \displaystyle\prod_{j=1}^{n-1} \frac{2R_j}{R_j+R_{j+1}}}{1+\displaystyle\sum_{i=1}^{n-1}\sum_{j=0}^{n-i-1}\rho_j\rho_{i+j}\epsilon^{\frac{-2Lis}{cn}}} \tag{5}$$

Equations 4 and 5 are evaluated for the one to four nominal impedance transforming exponential, Bessel, parabolic, and hyperbolic tapered transmission lines. The mathematical formulas for the surge-resistances of these lines can be found in Appendix I along with a plot of the surge resistances as a function of position on the line. Both equations 4 and 5 assume the form

$$V(s, L) = \frac{K\epsilon^{\frac{-sL}{c}}}{s\left[1+\displaystyle\sum_{j=1}^{99} n_i\epsilon^{\frac{-2Lis}{cn}}\right]} \tag{6}$$

where K = a constant
n_i = numerical coefficients

when numerical values of the various parameters are inserted. The response of the line $V(t,L)$ to a step input can be determined by taking the inverse Laplace transform of $V(s,L)$ given by equation 6. This operation can be simplified by expressing $V(s,L)$ in the form

$$V(s, L) = \frac{K}{s}\left[1-\sum_{j=1}^{\infty} m_i\epsilon^{\frac{-2siL}{cn}}\right]\epsilon^{\frac{-sL}{c}} \tag{7}$$

where m_i = numerical coefficient

The inverse Laplace transform of equation is given by

$$V(t, L) = K\left\{u\left[t-\frac{L}{c}\right] - \sum_{j=1}^{\infty} m_i u\left[t-\frac{2Li}{cn}-\frac{L}{c}\right]\right\} \tag{8}$$

where $u(t-\tau)$ = the unit step function

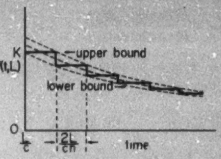


Fig. 2. Typical plot of V(t, L) versus t

For example, a typical plot of $V(t,L)$ versus t is illustrated in Fig. 2. The function in Fig. 2 that looks like a stair case is the transient on a step line. The curves which are dotted in are what would be expected as a result if the number of steps were infinite. Choice of a curve such as any of the dotted curves should insure a good approximation to the transient on a tapered line. Because the upper and lower curves can be thought of as upper and lower bounds on the transient solution of a tapered line, the curves plotted hereafter will not be stair case functions but will be instead smooth curves which are the means between the upper and lower bounds. Such a mean curve is shown in Fig. 2. This method of finding the transient response of a tapered line from the step line response yields the correct value of output at the time, $t=L/c$.

Response of lines which are matched nominally at the load and are fed by a signal generator whose internal impedance equals the complex conjugate of the nominal input impedance. The transient responses of these nominally matched lines are shown in Figs. 3(A) and 3(B). For the lines investigated the approximate results of equation 5 shown in Fig. 3(B) agree closely with the more exact results of equation 4 which are shown in Fig. 3(A). The two solutions are almost identical during the interval $1<tc/L<1.3$, but as tc/L increases the difference between solutions grows slightly larger and the approximate results appear to approach a slightly wrong final value as $tc/L\rightarrow\infty$. The more exact results, however, have a final value which is correct to eight significant figures. It can be seen from both Figs. 3(A) and (B) that the output of the exponential line decays more slowly during the interval $1<tc/L<1.65$ than the output of the other lines investigated. Of the lines investigated the hyperbolic line has the most rapidly fading transient. It also appears that the more negative the slope of the quantity $V(L/c,L)/V_0$ the quicker the transient dies away. However, no general conclusions can be drawn on the basis of only four different tapered lines and it is to be emphasized that everything stated thus far is known to be true only for the four cases investigated.

Response of lines which are nominally matched at the load and are fed by a signal generator of zero internal impedance. In the case of the lines just discussed the signal generator at the sending end consisted of a battery of voltage V_0, in series with a switch which is in series with an impedance having a value equal to the complex conjugate of the nominal impedance of the tapered line at the sending end. In this section the impedance which is in series with the ideal battery and switch is replaced by a short circuit. Fig. 4 shows the exact response on the lines considered. Under these terminations the behavior of the lines is quite different than it was in the previous section. In this case the hyperbolic line transient dies off least rapidly in the interval $1<tc/L<1.5$ and the exponential line is next best in that respect in the interval. In the smaller interval $1<tc/L<1.2$ the parabolic line transient decays more slowly than the exponential or Bessel line transients. Throughout the time intervals discussed the Bessel line's output is less than the output of the other lines. However, the Bessel line transient approaches its steady state with small oscillations whereas the hyperbolic line transient shows a marked degree of oscillation.

Conclusions

It can be seen that in the cases considered the exact and the approximate method of calculating transients on step lines yields very similar answers. This is particularly true for small values of tc/L where the results are almost identical. Thus, if it is desired to calculate only the initial part of a transient on a line the approximate method can be used to save time and effort.

Comparison of Figs. 3(A) and 4 indicates that the value of the internal impedance of the signal generator plays a very great role in the transient response of any given line. In most steady-state theories the line terminations have not been taken into account.

From the results obtained it appears that, with a given criterion in mind, for a fixed length of line and a fixed ratio of nominal impedance transformation there exists an infinity of optimum lines. More precisely, one optimum line might exist for each possible set of terminations. The companion paper established a criterion for the judgment of tapered lines, applied the criterion to a special case to find an optimum line, and considered the case of general terminations at both the input and the output.

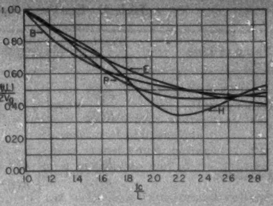
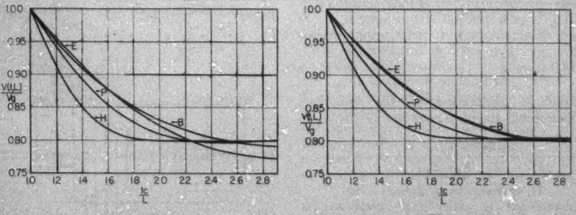


Fig. 4. Plot of output voltage versus time for the exponential E, Bessel B, parabolic P, and hyperbolic H lines. These lines have 1 to 4 ratio of impedance transformation, are excited by a signal generator of zero internal impedance, and are nominally matched at their receiving ends



(A)

(B)

Fig. 3. Plot of the output voltage versus time for the nominally matched 1 to 4 impedance transforming exponential E, Bessel B, parabolic P, and hyperbolic H lines in response to step input V_0 as calculated using the exact value of Z_{00}

Appendix I

Nomenclature of Tapered Transmission Lines

A transmission line having a characteristic impedance which varies with position along its length is called a tapered transmission line or a transmission-line-type pulse transformer. In the literature there are three or more distinct methods for naming such lines which often cause much confusion.[11] Many authors name their lines after the distributed inductance function. Examples of this are the exponential, hyperbolic, and parabolic lines. Some lines are known by the mathematical form of their logarithmic derivative of inductance function, e.g., the Gaussian line. A few lines are named after their own differential equation of voltage providing it is a well-known equation with a simple name. The Bessel line is named in that manner but is sometimes referred to as the linear line. In order to avoid any ambiguities which might arise from nomenclature, the tapered lines treated herein are tabulated in Table I along with their nominal characteristic impedance functions and their step-line approximations. (The nominal characteristic impedance Z_0 at a point on a line is the value of $\sqrt{l'/c'}$ one would calculate at that point from the physical dimensions of the line. A nominally matched line is terminated by a resistance which equals the nominal characteristic impedance of the line at the termination.)

Table I. Exact and Approximate Mathematical Forms of the Surge Resistances of the Lines Considered

Name of Line	Mathematical Form of R(x)	Step-Line Form of R(x)
Exponential	$\frac{R(x)}{R(0)}=\exp\left[\frac{x}{L}\ln\frac{R(L)}{R(0)}\right]$	$\frac{R_i}{R(0)}=\exp[1.40029(1-0.01i)]$ where $i=1, 2, 3\ldots 100$
Parabolic	$\frac{R(x)}{R(0)}=1+3\left[\frac{R(L)}{R(0)}-1\right]\frac{x^2}{L^2}-2\left[\frac{R(L)}{R(0)}-1\right]\frac{x^3}{L^3}$	$\frac{R_i}{R(0)}=1+9.0026828(1-0.01i)^2-6.0017885(1-0.01i)^3$
Hyperbolic	$\frac{R(x)}{R(0)}=\frac{1+\frac{R(L)}{R(0)}}{2}+\frac{1-\frac{R(L)}{R(0)}}{2}\tanh 6\left[\frac{x}{L}-\frac{1}{2}\right]$	$\frac{R_i}{R(0)}=2.5-1.5\tanh 6[1.0101010(1-0.01i)-0.5]$
Bessel	$\frac{R(x)}{R(0)}=1+\left[\frac{R(L)}{R(0)}-1\right]\frac{x}{L}$	$\frac{R_i}{R(0)}=1+3.0303030(1-0.01i)$

The Lines Used in This Paper

Let $x=0$ at the sending end and $x=L$ at the receiving end of the lines. Table I gives the name of the line, the general form of its characteristic impedance, and the exact numerical form of its characteristic impedance as used in this paper. Because only lossless lines are considered the characteristic impedance is sometimes called the surge resistance $R(x)$. Fig. 5 is a plot of the surge resistance of the step lines tabulated on Table I.

Appendix II

The Detailed Derivation of the Output Voltage of an N-Step Line

Consider the step line shown in Fig. 6. This line is composed of n sections of uniform equal length transmission line. The number i to the left of each section of line is the number of that section. The surge resistance of the ith section of line is R_i and Z_i is the impedance seen looking into the ith section. From ordinary transmission line theory

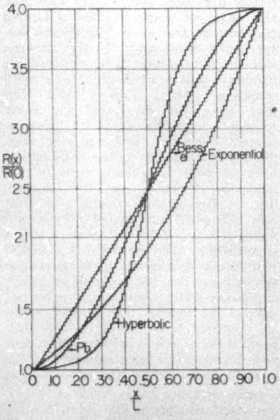


Fig. 5. Plot of the surge resistance versus position for the 1 to 4 impedance transforming Bessel, exponential, parabolic, and hyperbolic step lines

$$V_n=\frac{V_{n-1}}{2}\left\{\left[1+\frac{R_n}{Z_{n-1}}\right]\epsilon^{\theta}+\left[1-\frac{R_n}{Z_{n-1}}\right]\epsilon^{-\theta}\right\} \tag{9}$$

$$I_n=\frac{V_{n-1}}{2}\left\{\left[\frac{1}{Z_{n-1}}+\frac{1}{R_n}\right]\epsilon^{\theta}+\left[\frac{1}{Z_{n-1}}-\frac{1}{R_n}\right]\epsilon^{-\theta}\right\} \tag{10}$$

where

$\theta=s\Delta L/c$
$s=$complex frequency
$c=$the velocity of light
$\Delta L=$length of each section of line
$V_s=$the sending end voltage
$I_s=$the sending end current
$V_{n-1}=$the voltage impressed upon the $n-1$ section of line

Application of Kirchhoff's voltage law at the sending end yields

$$V_n+I_nZ_g=V_g \tag{11}$$

Upon substitution of V_n and I_n as given by equations 9 and 10 respectively equation 11 becomes

$$\frac{2V_n\epsilon^{-\theta}}{\left[1+\frac{Z_g}{R_n}+\frac{R_n+Z_g}{Z_{n-1}}\right]+\left[1-\frac{Z_g}{R_n}-\frac{R_n-Z_g}{Z_{n-1}}\right]\epsilon^{-2\theta}}=V_{n-1} \tag{12}$$

Repeated application of the recursion relationship stated in equation 12 yields

$$\frac{[2\epsilon^{-\theta}]^nV_g}{\left\{\left[1+\frac{Z_g}{R_n}+\frac{R_n+Z_g}{Z_{n-1}}\right]+\left[1-\frac{Z_g}{R_n}-\frac{R_n-Z_g}{Z_{n-1}}\right]\epsilon^{-2\theta}\right\}\times\prod_{i=1}^{n}\left\{1+\frac{R_i}{Z_{i-1}}+\left[1-\frac{R_i}{Z_{i-1}}\right]\epsilon^{-2\theta}\right\}}=V_0 \tag{13}$$

The product in the denominator of equation 13 causes that expression to be exceedingly unwieldy for $n>2$. However, if any two adjacent factors of the product are multiplied together great simplification follows. The product of all the factors so obtained is

$$\prod_{i=1}^{n-1}\left\{1+\frac{R_i}{Z_{i-1}}+\left[1-\frac{R_i}{Z_{i-1}}\right]\epsilon^{-2\theta}\right\}=\alpha_{n-1}\prod_{j=0}^{n-2}\frac{R_j+R_{j+1}}{R_j} \tag{14}$$

where α is defined in Appendix III. Substitution of equation 14 into equation 13 yields

$$V_0=\frac{2V_g\epsilon^{-n\theta}\frac{R_nR_{n-1}}{R_n+Z_g}\prod_{j=0}^{n-2}\frac{2R_j}{R_j+R_{j+1}}}{\alpha_{n-1}R_{n-1}[1+\rho_n\epsilon^{-2\theta}]+\beta_{n-1}R_n[1-\rho_n\epsilon^{-2\theta}]} \tag{15}$$

where

$$\rho_n\equiv\frac{R_n-Z_g}{R_n+Z_g},\ R_0\equiv Z_L,\ \alpha_0\equiv 1 \text{ (see Fig. 6)}$$

and β is defined in Appendix III.

Appendix III

By ordinary transmission-line theory the impedance seen looking into a length of uniform transmission line is

$$Z_{in}=\frac{1+\frac{Z_L-Z_0}{Z_L+Z_0}\chi}{1-\frac{Z_L-Z_0}{Z_L+Z_0}\chi}Z_0 \tag{16}$$

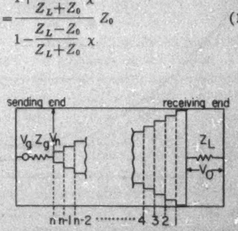


Fig. 6. Sketch of a typical step line which illustrates the method of denoting various sections of line

Vol
Pt.
195

where

ΔL = the length of the section of line
c = the velocity of light
s = complex frequency
Z_0 = the characteristic impedance of the section of line
Z_L = the load impedance
$\chi = \epsilon^{\frac{-2s\Delta L}{c}}$

The step line whose input impedance is being sought is composed of $n-1$ ordinary transmission lines connected end to end. Let the characteristic impedance of the ith line be denoted by R_i. Then for example, the characteristic impedance of the first line which is the one nearest the load and the $n-1$ the line are given by R_1 and R_{n-1} respectively. By application of equation 16 the impedance Z_1 seen looking into the first section of line is

$$Z_1 = \frac{1+\rho_0\chi}{1-\rho_0\chi} R_1 \qquad (17)$$

where

$$\rho_0 \equiv \frac{Z_L - R_1}{Z_L + R_1}$$

For convenience let

$$\rho_i \equiv \frac{R_i - R_{i+1}}{R_i + R_{i+1}}$$

The impedance Z_1 is the load of section 2. The impedance Z_2 seen looking into section 2 calculated by the use of equation 16 is

$$Z_2 = \frac{1+(\rho_1+\rho_0\rho_1)\chi + \rho_0\chi^2}{1-(\rho_1-\rho_0\rho_1)\chi - \rho_0\chi^2} R_2 \qquad (18)$$

Similarly

$$Z_3 = \frac{1+(\rho_2+\rho_0\rho_1+\rho_1\rho_2)\chi + (\rho_1+\rho_0\rho_2+\rho_0\rho_1\rho_2)\chi^2 + \rho_0\chi^3}{1-(\rho_2-\rho_0\rho_1-\rho_1\rho_2)\chi - (\rho_1-\rho_0\rho_2-\rho_0\rho_1\rho_2)\chi^2 - \rho_0\chi^3} R_3 \qquad (19)$$

For values of $n>3$, Z_n becomes extremely cumbersome. It would be most convenient if Z_n could be obtained without first calculating $Z_1, Z_2, \ldots Z_{n-1}, Z_n$. The knowledge of Z_{n-1} in literal form would enable the transient response of any step line to be exhibited in a form which would be quite useful for the purpose of generalization. An approximate literal value of Z_n has been determined by the process of ordinary induction. Inspection of equations 17, 18, 19, and many similar equations for impedance which are too lengthy to include here, indicates that the impedance Z_n will be of the form

$$Z_n = \frac{1+\sum_{i=1}^{n} \alpha_i\chi^i}{1-\sum_{i=1}^{n} a_i\chi^i} R_n \qquad (20)$$

The coefficient a_i and b_i appear to be of the form

$$a_i = \rho_{n-i} + \sum_{j=0}^{n-i-1} \rho_j\rho_{i+j} + \gamma_i \qquad (21)$$

$$b_i = \rho_{n-i} - \sum_{j=0}^{n-i-1} \rho_j\rho_{i+j} - \delta_i \qquad (22)$$

where γ_i and δ_i are third or higher order products of various combinations of the reflection coefficients $\rho_i(1 \leq i \leq n)$. In all cases worked out in this investigation, $\gamma_1 = \delta_1 = 0$. The other values of the coefficients γ_i and δ_i do not lend themselves to this inductive process. Because $\rho_i < 1$ the possibility exists that

$$|\gamma_i| \ll \left| \rho_{n-i} + \sum_{j=0}^{n-i-1} \rho_j\rho_{i+j} \right|$$

and

$$|\delta_i| \ll \left| \rho_{n-i} - \sum_{j=0}^{n-i-1} \rho_j\rho_{i+j} \right|$$

If the foregoing conditions are fulfilled the impedance Z_n is given by

$$Z_n = \frac{1+\sum_{i=1}^{n}\left[\rho_{n-i} + \sum_{j=0}^{n-i-1} \rho_j\rho_{j+i}\right]\chi^i}{1-\sum_{j=1}^{n}\left[\rho_{n-i} - \sum_{j=0}^{n-i-1} \rho_j\rho_{j+1}\right]\chi^i} R_n \qquad (23)$$

Equation 23 yields an approximate value of impedance for the step line. The exact value of impedance for a line of about 100 steps is indeed lengthy if written out in literal form. It is of the form

$$Z_n = \frac{\alpha_n}{\beta_n} R_n \qquad (24)$$

However, for any given tapered line it is a straightforward task to program an IBM *650* computer to calculate α_n and β_n by successive application of equation 16. The computer can also be programmed to find them by using equation 23. The IBM *650* was used in this manner to calculate the transients by both the exact and inexact method. The exact method used 2 hours of computing time and the approximate method used only one hour to find the complete transient solution to one line.

Nomenclature

a_i = coefficients of the power series in the numerator of Z_n
b_i = coefficients of the power series in the denominator of Z_n
c = velocity of light in the dielectric
c' = capacitance per unit length of line
ϵ = Euler's number
I_n = Laplace transformed current at the sending end of the step line
i = subscript of quantities pertaining to the ith section of uniform line
j = subscript of quantities pertaining to the jth section of uniform line
K = magnitude of the numerator of equation 5
l' = inductance per unit length of line
L = total length of the tapered line
n = number of steps used
n_i = numerical coefficients of the power series contained in equation 6
m_i = numerical coefficients of the power series contained in equation 7
O = subscript pertaining to the load end or receiving end of the tapered line
R_j = surge resistance of the jth section of uniform line
R_{n+1} = internal impedance of the signal generator
R_O = load impedance
s = complex frequency
t = time
$u(t-\tau)$ = unit step function ,which appears at $t=\tau$
V_i = Laplace transformed voltage at the ith section of uniform line
V_n = Laplace transformed voltage at the input or sending end of the tapered line
$V_O = V(s, L)$ = Laplace transformed voltage at the output or receiving end of the tapered line
$V_g(s, 0)$ = Laplace transformed voltage of the signal generator
$V(t, L)$ = output voltage of the line as a function of time
x = position along the tapered line measured from the sending end
Z_i = impedance seen looking toward the load into the ith section of uniform line
Z_g = internal impedance of the signal generator
Z_L = load impedance
$\alpha_0 = 1$
α_n = the numerator of equation 24
β_n = denominator of equation 24
γ_i = third and higher order products of reflection coefficients in numerator of Z_n
δ_i = third and higher order products of reflection coefficients in denominator of Z_n
ΔL = length of each section of uniform line ($\Delta L = L/n$)
$\theta = s\Delta L/c$
2ξ = denominator of equation 2
Π = standard symbol denoting products
ρ_i = nominal reflection coefficient at the ith section of uniform line
Σ = standard symbol denoting summation
τ = time delay
$\chi = \epsilon^{-2\theta}$

References

1. PROPAGATION DES COURANTS SINUSOIDAUX SUR DES LIGNES QUELCONQUES, C. Ravut. *Revue Générale de l'Electricité*, Paris, France, vol. 7, no. 19, 1920, pp. 611–15.

2. THE NONUNIFORM TRANSMISSION LINE, A. Starr. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 20, June 1932, pp. 1053–63.

3. MILLIMICROSECOND PULSE TECHNIQUES, I. A. D. Lewis, F. H. Wells. Pergamon Press Inc., New York, N. Y., vol. 5, 1954, pp. 63–93.

4. MICROWAVE TRANSMISSION, J. Slater. McGraw-Hill Book Company, Inc., New York, N. Y., 1942, pp. 69–75.

5. REFLECTIONS FROM SECTIONS OF TAPERED TRANSMISSION LINES AND WAVEGUIDES, N. Frank. *Report no. 17*, Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., section 43, Jan. 1943.

6. FOURIER TRANSFORMS IN THE THEORY OF INHOMOGENEOUS TRANSMISSION LINES, F. Bolinder. *Transactions*, Royal Institute of Technology, Stockholm, Sweden, no. 48, 1951.

7. THE HYPERBOLIC TRANSMISSION LINE AS A MATCHING SECTION, H. J. Scott. *Proceedings*, Institute of Radio Engineers, vol. 41, 1953, pp. 1654–57.

8. PARABOLIC TRANSMISSION LINE, R. Yang. *Ibid.*, vol. 43, 1955, p. 1010.

9. A TRANSMISSION LINE TAPER OF IMPROVED DESIGN, R. W. Klopfenstein. *Ibid.*, vol. 44, 1956, pp. 31–5.

10. THE OPTIMUM TAPERED TRANSMISSION LINE

MATCHING SECTION, **R. E. Collin.** *Ibid.*, vol. 44, 1956, pp. 539–57.

11. THE OPTIMUM TAPERED TRANSMISSION LINE MATCHING SECTION, **F. Bolinder, R. E. Collin.** *Ibid.*, vol. 44, pp. 1055–56.

12. COMPUTATION OF THE IMPEDANCES OF NONUNIFORM LINES BY A DIRECT METHOD, **Louis A. Pipes.** *AIEE Transactions*, vol. 75, pt. I, Nov. 1956, pp. 551–54.

13. PULSE TRANSIENTS IN EXPONENTIAL TRANSMISSION LINES, **E. R. Schatz, E. M. Williams.** *Proceedings*, Institute of Radio Engineers, vol. 38, 1950, pp. 1208–12.

14. DESIGN OF EXPONENTIAL LINE PULSE TRANSFORMERS, **E. M. Williams, E. R. Schatz.** *Ibid.*, vol. 39, 1951, pp. 84–6.

15. ENGINEERING PROBLEMS IN THE EXPONENTIAL TRANSMISSION LINE PULSE TRANSFORMER DESIGN, **J. Kukel.** *Doctoral Dissertation*, Carnegie Institute of Technology, Pittsburgh, Pa., May 1953.

16. I TRANSITORI NELLA LINEE NON UNIFORMI, **G. Zin.** *Alta Frequenza*, Milan, Italy, vol. 10, 1941, pp. 707–50.

17. THE OPTIMUM TRANSMISSION-LINE PULSE TRANSFORMER, **F. J. Young, E. R. Schatz, J. G. Woodford.** *AIEE Transactions*, see pp. 220–23 of this issue.

◆

Discussion

L. O. Barthold (General Electric Company, Schenectady, N. Y.): The authors have done a very thorough job in their analysis of the complex problem of nonuniform transmission lines. An approximate solution of the same problem was recently reported in the literature.[1] The latter solution was made in conjunction with investigations of the response of transmission towers to lightning strokes.

In the approximate solution referred to, the line is considered first as a stepped line, but as opposed to the authors' assumption, the steps are taken in equal impedance ratios instead of equal electrical distances. This results in the same reflection and refraction coefficients at each junction.

The solution was made from a generalized lattice diagram in which the number of steps was made to approach infinity. The important approximation made was in neglecting all waves which appear with reflection coefficients of second and higher orders.

For the case where the nominal surge resistance of the line is matched at both ends, the above assumption leads to the general expression

$$\frac{V(t,0)}{V_s}=1+\frac{1}{2}\log_e\frac{R\left(\dfrac{tc}{2L}\right)}{R(0)}$$

Since the reflected terms at the sending end appear with only odd powers of the reflection coefficient, the terms which are omitted here are third order and higher and the approximation is a reasonably good one.

Given the nominal surge resistance as a function of travel time, the foregoing solution can be made very quickly by hand calculation.

The same analysis applied to receiving end voltage omits second and higher order terms which, as the authors show, are quite significant. The foregoing approximation is correct in this case only for the initial value of the receiving-end voltage.

Since the assumption of equal impedance ratio steps greatly simplifies the generalized lattice approach, do the authors feel that this assumption might simplify the solution they have made?

Also, since the ultimate solution made by the authors is the summation of a finite series of uniform line segments, it would seem that an equally good approximation could be obtained by representing the line as a lumped parameter circuit. It would be interesting to know if the authors considered this approach.

REFERENCE

1. AN APPROXIMATE TRANSIENT SOLUTION OF THE TAPERED TRANSMISSION LINE, **L. O. Barthold.** *AIEE Transactions*, vol. 77, pt. III, 1957 (Feb. 1958 section), pp. 1556–61.

F. J. Young, E. R. Schatz, J. B. Woodford: The authors are grateful for Mr. Barthold's comments. It is gratifying to know that his theory based on equal steps of impedance checks with the initial values of the output voltage calculated by our equations. The authors briefly considered using such an approach but because it did not permit mathematical generalization as easily as the equal length method it was abandoned. Thus, in the authors' opinion the equal impedance steps would not have simplified the solution.

Representation of a transmission line by lumped parameters is not a good approximation at all frequencies. Since transients are being investigated and therefore high-frequency components, an exorbitantly large number of lumped elements might be necessary to obtain an adequate approximation, perhaps more than the number of transmission-line elements finally chosen for this work.

A Single-Stator Meter for Two Phases of a 4-Wire Y

E. W. SCHWARZ
MEMBER AIEE

Synopsis: By modification of a 3-wire single-phase meter, a single-stator watt-hour meter has been developed for the measurement of two phases of a 4-wire Y 3-phase distribution system. The paper discusses the modifications of the meter together with its performance characteristics.

WHEREVER a 4-wire Y 3-phase distribution system is in use and from which energy is derived by means of two phases and the neutral, there arises the problem of metering the 3-wire system, so-called "network service." The usual method of metering the 3-wire network system is by means of a 2-stator polyphase meter, which essentially consists of two electromagnetic structures, each of which has a voltage and current coil. While present-day production and design techniques have reduced the cost of the 2-stator network meter, there still remains a considerable price differential between the network meter and a 3-wire singlephase meter. It has been the desire of utilities and meter manufacturers to provide a watt-hour meter that would be applicable for network metering, yet substantially less in cost than that of a 2-stator polyphase meter.

A single-stator meter has been developed and is being manufactured that will measure the energy of a network system. Its cost is considerably less than that of a 2-stator network meter.

While there is a definite price advantage in the use of single-stator network meters, utilities interested in their possibilities should not overlook certain disadvantages. They will have to make the final decision as to the importance of cost savings versus usage and testing disadvantages.

The Current Electromagnet Structure

As previously stated, the design of the single-stator meter (Fig. 1) was developed with the primary purpose of

---

Paper 59-17, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 14, 1958; made available for printing November 13, 1958.

E. W. SCHWARZ is with the Sangamo Electric Company, Springfield, Ill.

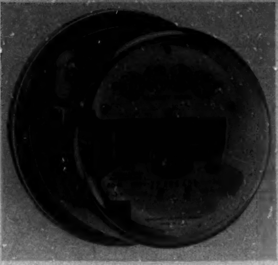


Fig. 1. Single-stator network meter

effecting the maximum possible cost reduction from that of the 2-stator network meter. Consequently, the decision was made to use the single-stator 3-wire single-phase meter as the building block for the single-stator network meter. There are two possible choices in converting the single-stator construction into a network meter. One choice consists of using phase-to-phase voltage for the voltage coil, which dictates changing the phase of both line currents inside the meter. While this scheme was considered, it was felt that the addition of two resistors and two reactors necessary to accomplish the aforementioned phase shifting was not advisable in order to achieve the maximum possible cost reduction.

The other possible solution, and the one that has achieved the major objective, is the use of phase-to-neutral voltage for the voltage coil, resulting in the requirement that only one line current need be shifted in phase. For this scheme only one resistor and one leakage core are required to convert the single-stator 3-wire single-phase construction for use on network service.

The current electromagnetic structure is illustrated in Fig. 2. There are two current coils wound on a common laminated core. The upper coil is a 2-turn coil connected in line A as shown in Appendix I. The lower coil is an 8-turn coil connected in line B. As illustrated in Appendix I, the current I_C through the 8-turn coil must lag the line current I_B by 60 degrees, which implies, since I_B lags the voltage E, that the current I_C must lag the voltage E by some angle greater than 60 degrees. To accomplish the necessary phase shift of I_C, an 8-turn coil was chosen, which increases the ratio of the inductive reactance to resistance of the coil. In addition, a laminated leakage core, so designed as to be movable with respect to the main laminated core, is located between the 2- and 8-turn coils. Not only must the phase of I_C be precisely adjusted, but the magnitude of I_C must be such so that for comparable line currents the torque-producing flux from the 8-turn coil has the same magnitude as the torque-producing flux from the 2-turn coil. Stated another way, the magnitude of I_C must also be adjustable in addition to its phase. In order to complete the adjustment of magnitude and phase of I_C, a manganin resistor is riveted and brazed across the 8-turn coil. The resistance of this shunt is adjusted by removing material. To summarize the foregoing statements, it is possible to obtain correct phase and magnitude of the current in the 8-turn coil by adjusting the position of the leakage core and changing the resistance of the resistor. These two adjustments are preset in the manufacturing process and need not be altered.

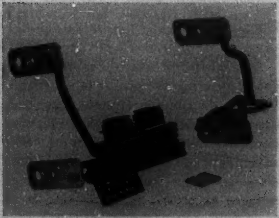


Fig. 2. Current electromagnetic assembly

Pole-tip extensions are placed between the poles of the current core structure in order to obtain additional torque. In this manner the effective pole area of the current core is increased, resulting in increased driving torque.

All current carrying parts of the current electromagnetic structure have been designed to carry 60 amperes continuously while still maintaining a tolerable temperature rise inside the meter. The meter is therefore rated as a class *60* meter. Since the thermal characteristics limit the maximum current to 60 amperes, and as a result of using the same slow disk speed of a single-phase meter with a $K_h=6$, it was determined that no overload compensation was necessary, a factor again favorable in reducing the cost.

The Voltage Electromagnetic Structure

In Fig. 3 the various components of the voltage electromagnetic assembly are shown. The number of turns and wire size of the voltage coil were changed from that of a standard 120-volt single-phase coil in order to obtain additional driving torque. Even so, the watts loss in the potential coil is not excessive and, in fact, is 25% less than the watts loss of a comparable 2-stator meter. The power-factor adjustment consists of a phasing band, the resistance of which is altered in order to adjust the meter for lagging power-factor loads. This adjustment is also preset in manufacture and eliminates the need for any counter compensation of the current electromagnetic core. Class II temperature compensation is accomplished by the addition of temperature-sensitive nickel-steel to the phasing-band assembly. The light-load adjustment is the same as that used for certain single-phase meters, and consists of a movable vane, the position of which compensates any light-load inaccuracies.

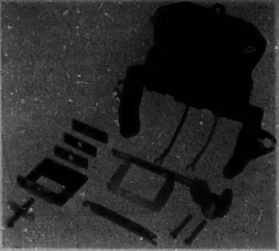


Fig. 3. Voltage electromagnetic assembly

Only the light-load and full-load adjustments are available for future recalibration. The coil-balance and power-factor adjustments having been fixed during manufacture.

The Phase-Sequence Indicator

In Appendix I, the necessity of correct phase sequence for proper operation of single-stator network meters becomes apparent. To insure that at the time of installation the phase sequence is correct, a phase-sequence indicator is supplied as an integral part of the meter. The indicator consists of two resistors, a capacitor, and a neon bulb, all encapsulated in an epoxy resin. When incorrect phase sequence is applied to the meter, the neon bulb will light. When correct phase sequence is applied, the bulb will not light. In this manner, correct installation is assured, and if at any time the phase sequence is reversed, an alarm is readily available.

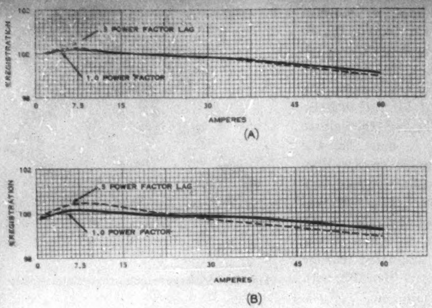


Fig. 4 (above). Load curves

A—Line-to-line
B—Line-to-neutral
Type JNS watt-hour meter

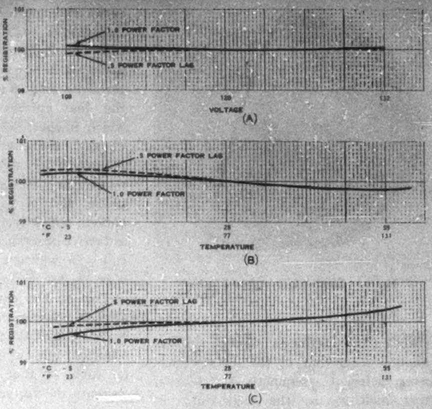


Fig. 5 (right). Voltage and temperature variation curves

A—Voltage characteristic at rated current
B—Temperature characteristic at rated watts
C—Temperature characteristic at 10% rated watts

Testing Procedures

The calibration and testing of single-stator network meters is not as straightforward as a single-phase meter. The basic difficulty is that of assuring balanced voltages in the testing procedures. Because of the fact that there are many possible ways of testing, some of which may be more applicable to certain utilities than others, this discussion will only describe one possible procedure.

Appendix II illustrates a test circuit applicable for use in testing the single-stator meter for line-to-line loading. If the distribution system is well balanced (magnitude and phase), it is not necessary to resort to phase-shifting networks. If there is some doubt as to the degree of balance of the voltages, it is not too difficult to construct very accurate phase-shifting networks that eliminate the need for maintaining voltage balance. Such networks operate from the 120 volts that would be applied to the potential coil of the meter under test, and produce a voltage having a magnitude of 208 volts which leads the 120 volts by 30 degrees. In this manner the rotating standard operates at unity power factor when a unity power-factor load is applied to the meter. A 50% lagging power-factor test can be made by the usual practice of cross-phasing the current.

Table I. Characteristic Data of Single-Stator Network Meter

Rated watts	3,600
K_h, watt-hours/revolution	6.0
Maximum current, amperes	60
Torque at rated watts, millimeter-gram	40.0
Potential Circuit Data at 120 volts:	
Loss, watts	1.5
Potential exciting current, amperes	0.093
Current Circuit Data at 15 amperes:	
Loss, watts	0.28
Volt amperes	0.29

For the balancing adjustments between the two coils, the registration of the 2-turn coil is compared with that of the 8-turn coil at both unity and 50% lagging power factor. The registration of the 2-turn coil is easily obtained by a standard single-phase test. A scheme similar to that shown in Appendix II can be used for obtaining the registration of the 8-turn coil. In this case the phase-shifting network must be designed to produce a voltage that is shifted 60 degrees from the input voltage. By the use of such phase-shifting networks there is no need for maintaining a balanced 3-phase supply at any time in the calibration procedure. The foregoing procedure guarantees correct calibration for line-to-neutral and line-to-line loading.

The simplest field test that could be used would consist of testing the 2-turn coil. This coil, since it does not shift the phase angle of the line current, can be tested by a standard single-phase test and will indicate possible deviations in the other calibrations due to the common-core structure used in the single-stator construction.

As mentioned previously, it is usually advisable because of possible differences between utility practices and state codes to discuss the type of test more appropriate for the user.

Meter Accuracy

While discussions of construction and design details of watt-hour meters may be of academic and practical interest to utilities, the ultimate criteria for acceptance is, of course, the degree of accuracy with which a particular design will indicate the true energy delivered to a load.

Fig. 4 indicates the average load curve accuracy of the meter. Since both line-to-line and line-to-neutral loading is important, separate curves are shown for both cases. Voltage variation and temperature variation accuracies are shown in Fig. 5. While it is seen that the accuracy curves are not quite as favorable as those for a 2-stator polyphase meter, it is felt that they are of sufficient accuracy for present-day metering practices.

In Table I are listed miscellaneous data for the single-stator meter.

Conclusions

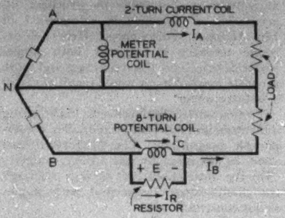
A single-stator watt-hour meter for use in metering two phases of a 4-wire Y 3-phase distribution system has been described wherein the primary objective has been that of cost reduction. The advantages and disadvantages have been discussed along with the major design changes necessary in converting a 3-wire single-phase meter into a single-stator network meter. The accuracies,

Vol.
Pt. 1
1959

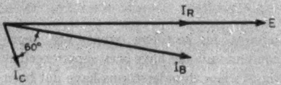
while not identical with those of a multi-stator meter, are well within commonly accepted standards.

Appendix I

Discussion of 60-Degree Phase Shift



E_{BN} lags E_{AN} by 120 degrees



It is possible to demonstrate why the current I_C must lag the line current I_B by 60 degrees by considering the following expressions.

If $E_{AN}(I_A)$ is defined as the dot product of the two vectors E_{AN} and I_A, the power P to the load then becomes

$$P=E_{AN}(I_A)+E_{BN}(I_B)$$

For balanced voltages and ABC phase sequence

$$E_{BN}=E_{AN}\epsilon^{-j120}$$

therefore

$$P=E_{AN}(I_A)+E_{AN}\epsilon^{-j120}(I_B)$$

Now if I_C is chosen such that

$$I_B=KI_C\epsilon^{j60}$$

where K is a positive real constant

$$P=E_{AN}(I_A)+E_{AN}\epsilon^{-j120}(KI_C\epsilon^{j60})$$

which can be expressed as

$$P=E_{AN}(I_A)+E_{AN}\epsilon^{-j120}(KI_C\epsilon^{-j120}\epsilon^{j180})$$

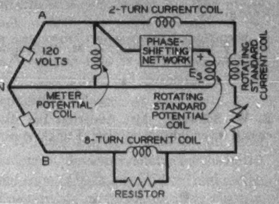
or

$$P=E_{AN}(I_A)+E_{AN}(-KI_C) \text{ qed}$$

The constant K is accounted for in the design of the meter by the proper choice of the number of turns of the current coil and value of the shunting resistor. The minus sign merely indicates that the winding direction of the coil must be reversed from that direction which would normally be used.

Appendix II

One Possible Test Procedure for Checking Type JNS Meter for Line-to-Line Loading



The phase shifting network is designed such that the magnitude of E_S equals 208 volts and such that the phase of E_S leads E_{AN} by 30 degrees.

◆

Discussion

J. Anderson (General Electric Company, Somersworth, N. H.): This paper is an excellent addition to the literature on watt-hour meters. The discusser would like to comment on the author's pertinent advice to utilities that they weigh certain disadvantages of single-stator network meters and then decide for themselves the importance of cost savings versus usage and testing disadvantages. It should be emphasized that the significance of these disadvantages is dependent on the size of the utility, the quantity of network meters being used, and the characteristics of the power distribution system. Thus these factors are not of any major significance for large urban utilities using large quantities of network meters and having a distribution system with a high degree of voltage and frequency stability.

The major use disadvantage is the necessity of obtaining the correct phase sequence and applies to any single-stator network meter. Color-coded wiring, together with the proper wiring specifications, is being effectively used by utilities to overcome this difficulty.

A second major disadvantage is the difficulty of testing. The meter described in this paper cannot be tested with standard single-phase meter test equipment and procedure. This is because this meter cannot be tested with the current coils connected in series and with a voltage applied which is in phase with the load current. Appendix II points out one possible test procedure for line-to-line loading which is equivalent to a normal single phase meter test procedure. However, this procedure requires special test equipment having a phase shifting network to supply potential excitation of the proper magnitude and phase angle to the potential coil of the watt hour standard.

The alternate design with 2-phase shifting networks, mentioned by Mr. Schwarz at the outset of his paper, does not have this testing disadvantage. It can be shop or field tested by utilities with exactly the same equipment and procedure as they are presently using for single-phase 3-wire meters. This can be readily understood by recognizing that when this meter measures a line-to-line load on a normal network service the voltage applied to the potential coil of the meter is the voltage applied to the load. The normal 3-wire single-phase meter test applies voltage and current which are in phase. Utilities using very large numbers of network meters may be able to justify special test equipment; however, the ability to test a single-stator network meter with existing equipment would appear to be highly desirable for those utilities using only a moderate number of these meters. It is of course necessary in the manufacture and factory calibration of either type of network meter to have accurately balanced voltages with precise control of both magnitude and phase angle, which is very practical for a manufacturer to achieve with special test equipment.

A further design and application consideration for single-stator network meters involves evaluating the effect of voltage unbalance. The alternative design of network meter utilizes the line-to-line voltage which is the vector sum of the two line-to-neutral voltages. This meter was described in a paper by McQuarrie.[1] The voltage error seen by this meter is comparable with that of the standard 3-wire single phase meter which also utilizes the sum of the line-to-neutral voltages. The meter design described by Mr. Schwarz utilizes only one of the two line-to-neutral voltages and therefore is subject to relatively larger errors with voltage unbalance at higher power factors as compared with the alternative single-stator network meter.

A tabulation of comparative voltage errors is given in Table II.

Two final considerations that utilities should weigh are performance with varying frequency and power factor. Single-stator network meters in general are more frequency sensitive than the conventional 2-stator network meters. While the meter described in this paper is considerably more frequency sensitive than the alternative design for line-to-line loading (208-volt

Table II. Comparative Voltage Errors

Power Factor	Meter Described in Paper; Calculated Voltage Error, %	Alternative Design; Calculated Voltage Error, %
1.0% voltage unbalance		
Balanced line-neutral load		
1.0	±0.50	±0.01
0.8	±0.50	±0.22
0.5	±0.50	±0.50
Line-line load		
1.0	±0.50	0
0.8	±0.28	0
0.5	±0.01	0
2.5% voltage unbalance		
Balanced line-neutral load		
1.0	±1.25	±0.02
0.8	±1.25	±0.50
0.5	±1.25	±1.25
Line-line load		
1.0	±1.25	0
0.8	±0.75	0
0.5	0	0

loads), these errors probably are not significant when used on the closely controlled frequency of almost all of today's modern power systems.

In general, the low power factor performance of single-stator network meters is significantly poorer than that of 2-stator meters. However, it does meet the requirements of the Code for Electricity Meters, and it is probable that the accuracy obtained, when these meters are used on the normal network loads having reasonably good power factors, will be considered satisfactory.

REFERENCE

1. Watt-Hour Meter for Two Phases of a 4-Wire Y 3-Phase System, A. M. McQuarrie. *AIEE Transactions*, vol. 77, pt. I, May 1958, pp. 152–56.

E. W. Schwarz: Mr. Anderson has discussed several disadvantages of single-stator network meters, two of which will be commented upon in this closure.

One of the difficulties mentioned is that of testing. The alternative design mentioned by Mr. Anderson also requires special test equipment whenever a check is desired for balance between the two current circuits. This is because both line currents are shifted in phase internally by the use of two phase-shifting networks. The design described in the paper also requires special equipment or balanced voltages when testing the meter with its current circuits in series. It should be realized that one of the current circuits can be checked (and also used) as a 2-wire single-phase meter. Consequently, it seems as if both designs have their own testing difficulties.

Regarding the comments pertaining to errors introduced by voltage unbalance, the particular types of voltage unbalance and associated errors as described do exist in single-stator meters. The author might mention that there exist other types of voltage unbalance, particularly on unbalanced loads, wherein the alternative design would have comparable errors and wherein the described design would have no error.

Once again, without resorting to figures, both designs are subject to errors introduced by voltage unbalance. Even so, the probable errors that might be encountered by such unbalances have the same order of magnitude as the inherent overload inaccuracies as described in reference 1 of Mr. Anderson's discussion. Hence, if extreme accuracy is required, both single-stator designs have disadvantages and more accurate measurement would result by using a 2-stator meter.

To summarize, both schemes have testing difficulties and inherent inaccuracies of comparable magnitude. The final conclusion must be obtained by the utility, namely, "Do the anticipated cost savings outweigh the disadvantages associated with single-stator construction?"

Electrical Protection for Transistorized Equipment

D. W. BODLE
MEMBER AIEE

J. B. HAYS, JR.
MEMBER AIEE

THE TREND TO miniaturization of components and the expanding use of transistor circuitry in communication equipment has necessitated the development of new protection measures capable of limiting foreign potentials and currents to values much lower than has heretofore been necessary.

Older conventional equipment was designed to co-ordinate with protector discharge gaps having a minimum nominal gap spacing of 0.0028 inch and a sparkover value of 500 peak volts. In order to insure an adequate margin above the sparkover value of the protectors, such equipment was required to withstand a minimum dielectric test of 500 volts rms. In the case of transistorized equipment, however, it is usually not feasible to attain such a withstand level. It has been necessary to develop additional protection measures capable in some instances of limiting potentials to only a few volts.

This paper discusses the source and nature of the foreign potentials hazardous to the dependable operation of communication equipment and the additional protection measures that have been successfully employed to protect relatively delicate equipment of recent design.

Sources of Foreign Potentials

Communication equipment associated with land-transmission facilities both aerial and buried is frequently exposed to abnormal potentials produced by lightning; this, in practice, constitutes the major electrical hazard. Although communication lines and cables frequently occupy pole facilities jointly with power circuits, because of good construction and co-ordination practices, power contacts are relatively few and equipment damage from this source is much less likely than from lightning.

Equipment powered from aerial power distribution facilities in suburban and rural areas is vulnerable to lightning surges entering an installation over the service conductors. Furthermore, failures in rectifiers employing semiconductor diodes have been experienced at times when thunderstorms have not been reported in the general area.

For economic reasons, such equipment is frequently designed to have a relatively small inverse-voltage safety factor; consequently, it has become necessary to consider the effects of switching and other operating transients from the power source. Furthermore, although not within the intended scope of this paper, it is interesting to note that semiconductor diodes employed in power rectifiers, in addition to their vulnerability to foreign potentials, may also be susceptible to damage from internal transients that develop when normal line voltage is applied. This situation should be anticipated by the designer and has recently received some mention in manufacturers' bulletins and published papers.[1]

The exposure of a radio installation to lightning strokes to the antenna is well recognized, and adequate consideration is generally given to the protection of an installation against damage from this source.[2] However, it is not so generally recognized that the probable incidence of strokes to an antenna may in many cases average less than one a year, yet the station equipment will probably be exposed each year to several lightning surges from the power line. Some of these surges may be of sufficient magnitude to cause tower lamp burn out, relay contact burning, and rectifier troubles.

Since the measures employed to protect equipment from foreign potentials on the communication transmission facilities are considerably different from those em-

---

Paper 59-71, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, February 1–6, 1959. Manuscript submitted October 31, 1958; made available for printing November 28, 1958.

D. W. Bodle and J. B. Hays, Jr., are with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

The authors wish to acknowledge the assistance of K. D. Smith, H. S. Veloric, and other members of the Bell Telephone Laboratories Semiconductor Staff.

Vol.
Pt. 1
1959

ployed on the associated power circuits, it will facilitate the discussion to consider them separately.

COMMUNICATION LINES AND CABLES

Lightning surges of sufficient magnitude to be significant from the protection standpoint appear in communication lines and cables as a result of the following.

1. Direct strokes to aerial and buried plant.
2. Exposure of buried plant to earth potential gradients.
3. Exposure of aerial plant to flashover from adjacent facilities and potential rise of associated grounds.
4. Induction in nonshielded facilities.

EXTERNAL POWER FACILITIES

Several years ago rectifier tube and capacitor failures were experienced in power supplies at intermediate telephone-carrier and radio-repeater stations normally receiving their a-c power over aerial distribution lines. Such installations are usually in rural locations and served by small lines having few other users in the immediate area. Lightning surges from the power lines were suspected of being the probable cause of the equipment damage. Although little was or has been done to determine precisely the incidence and characteristics of such surges, the effectiveness of the applied protection measures described later in this paper definitely established the power lines as the source of the trouble. It was incidentally found that after protection was applied to the power service, there was a marked reduction in other equipment troubles, such as lamp burn out in tower warning lights. Rectifiers employing semiconductor diodes may be even more vulnerable to power-line surges than thermionic rectifiers. It is advisable therefore to consider the protection requirements of such equipment during the design stage.

Nature of Lightning Potentials on Communication Lines and Cables

Lightning potentials in communication plant are primarily longitudinal (conductor-to-earth), but metallic voltages (conductor-to-conductor) develop because of circuit unbalances as a result of insulation sparkover or dissimilar operation of some types of protection devices. When there is no circuit path to ground, or such a path is interrupted by a transformer, it is not difficult to provide a minimum insulation withstand level with respect to ground of 700 peak volts; 0.0028-inch protector gaps will in such cases provide adequate protection against longitudinal voltages. However, in circuits having direct paths for the line conductors to ground through components, there may be a problem concerning the ability of these components to conduct currents produced by longitudinal voltages less than the sparkover value of the protectors. In these instances it is frequently possible, by simple design changes, to bring the withstand ability of critical components up to the level where they will be adequately protected by the discharge gaps.

In the more general situation, coupling transformers are used and the circuit elements on the secondary side are exposed only to metallic voltages; however such voltages may reach magnitudes equal to the sparkover value of the higher of the two protector gaps on the primary side of the coupling transformer (500 to 600 peak volts). For low-voltage circuit components which may be damaged by voltages as low as 40 to 50 peak volts, as in the case of transistors, additional protection measures are required.

A lightning surge at or near the stroke point is essentially unidirectional. Such a surge will, however, be modified both in magnitude and waveshape as it propagates through communication lines and cables; consequently it is impossible to anticipate the nature of such surges except in a very general way. Observations on both open wire and cable plant indicate that surges range in waveshape from simple impulses to complex oscillatory transients. These transients vary greatly with the type of plant involved and the operation of associated protection devices. Surges that arrive at a terminal point on a line from a stroke some distance away, or that have otherwise been attenuated so that they will not spark over protector blocks, do not have frequency components as high as commonly supposed. For example, some observations of surges on an open-wire toll line indicated varying amplitude components in the order of 20 kc having durations of 1.0 to 1.5 milliseconds. Metallic transients observed in a loaded cable pair were highly damped and had fundamentals of 1,500 to 2,000 cycles with amplitudes decaying to half value in a millisecond or less.

It was from such data that a 10x600-μsec impulse test wave* was established as suitable for laboratory testing of protection diodes on the basis that such a wave produces equivalent duty on the test specimens. Such a test wave is relatively easy to produce in the laboratory and, through correlation with field experience secured with equipment so tested, provides a satisfactory basis for evaluation of circuits and components.

In one study of foreign metallic voltages on a pair in a metallic sheath cable, 249 surges ranging in amplitude from 40 to 550 peak volts were observed during a total of 10 thunderstorms. The magnitude distribution of these metallic surge voltages is shown in Fig. 1. The upper limit of 550 peak volts was due to the presence of 0.0028-inch protector discharge gaps.

Protection Measures

Protector discharge gaps are the most economical devices available for the initial limitation of both longitudinal and metallic potentials in communication circuits. Such protectors, having nominal gap spacings of 0.0028 inch, are used at subscriber and terminal equipment points with a reasonable amount of protector maintenance. Smaller gap spacings are impractical from a maintenance standpoint; where protection is required below about 600 peak volts, some auxiliary method must be employed. Semiconductor protection diodes have been successfully employed in this lower range. Generally it has been possible to protect transistorized equipment using available voltage-limiting diodes in the range from

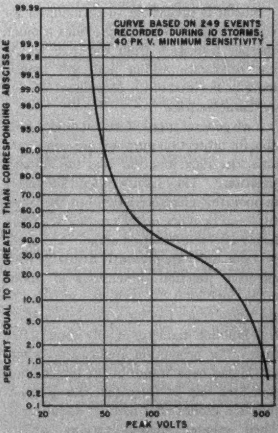


Fig. 1. Distribution of metallic voltage surges observed on a loaded cable pair during local thunderstorms

---

* This is a unidirectional wave that rises to crest value in 10 μsec and decays to 1/2 of its crest value in 600 μsec measured from $t=0$.

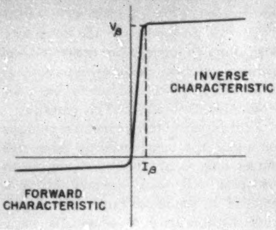


Fig. 2. Generalized semiconductor diode characteristic

about 6 to 68 peak volts. Specific methods of application will be discussed later.

Semiconductor Protection Diodes

General Characteristics

Semiconductor diodes have inverse conduction characteristics that make them exceptionally fine voltage limiters. Such devices present a very high resistance when inverse voltages below a critical "avalanche breakdown" value are applied, and also show a well-defined region of substantially constant voltage limitation above this critical breakdown value; see Fig. 2. Although the term "breakdown" is commonly used in referring to the critical voltage range in which a junction changes rapidly from a high-resistance barrier to a good conductor, the term may be somewhat misleading as the phenomenon is not destructive when confined to design limits.

Silicon is preferable to germanium for voltage limiters since it provides a much sharper transition from the high to the low resistance region and will handle more power without a permanent change in characteristics.

Two general types of silicon diodes suitable for protection purposes are now available, the alloy junction[3] and the diffused junction.[4] They have a general similarity in operating characteristics, but the latter type can be more easily produced, having larger effective junction areas. Diffused-junction units are presently available with effective junction dimensions of about 0.125 inch by 0.125 inch, as compared to the alloy types having a maximum effective junction diameter of only about 0.025 inch. For protection applications where the higher capacitance of the larger junction area can be tolerated, the diffused-junction diodes offer distinct advantages. Their greater current-carrying ability and their lower effective resistance result in a smaller increase in voltage above the breakdown value, with increasing surge current as may be noted in Fig. 3. In the application of protection diodes to circuits where surges of both polarities may appear, it is necessary to use two units in series and oppositely poled.

The forward conductivity characteristics of semiconductor junctions may also have application for protection purposes in situations where the normal circuit-voltage levels are very low. A significant change in forward junction resistance occurs in the case of silicon at about 0.5 volt. Although the transition is less sharp than in the inverse direction it may provide sufficiently effective voltage limitation, in spite of the relative softness of the voltage characteristic, because of the very low magnitude at which it starts to limit.

To utilize the forward characteristic in protection applications, it is necessary to employ two oppositely poled junctions in parallel. Such double-unit devices, encapsulated in one envelope similar in appearance to a resistor, have been developed. Because of the nature of the junction material employed there is every indication that they will be very low in cost when in production. In spite of their small physical size, such devices have the ability to conduct surprisingly large values of surge current and have interesting circuit-application possibilities which will be discussed later.

Inverse-Voltage Breakdown Ratings

Protection diodes are rated on the basis of their nominal inverse breakdown value. This rating refers to the small region at the knee of the inverse-voltage current-characteristic curve and is established on the basis of a specific current value, commonly 0.5 milliampere. At present,

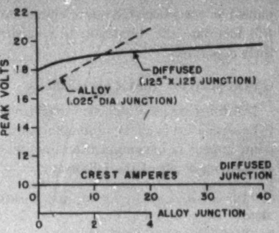


Fig. 3. Voltage-limiting characteristic of typical silicon protection diodes

practical protection diodes can be produced having breakdown ratings ranging from about 6 to 150 volts, direct current.

The response time of a semiconductor diode is extremely fast, in the order of milli μsec. It may therefore be assumed that for protection purposes, the inverse-voltage breakdown point for a lightning surge will be the same as the d-c rated value.

Inverse-Surge-Current Carrying Ability

Semiconductor diodes have a higher resistance in the inverse conducting direction than in the forward direction. Since in most protection applications the diodes are required to conduct current in either direction, their inverse-current carrying ability is the restricting factor in such applications.

Data on inverse current, also referred to as reverse current, given in manufacturers' data sheets generally refer to steady-state direct current and consequently have little relationship to the ability of a diode safely to conduct inter-

Table I. Inverse-Surge-Current Carrying Ability of Silicon Protection Diodes*

Type of Diode	Diameter of Junction, Inches†	Wafer Size‡	Average Breakdown Voltage	Waveshape of Applied Surge, μsec	Withstand Currents (Crest Amperes) Minimum	Median	Maximum
Silicon alloy	0.025		6.0	$1\frac{1}{2}$x40	80.0	90.0	100
			6.0	10x600	10.0	12.0	24
			9.5	$1\frac{1}{2}$x40	11.0	17.0	55
			9.5	10x600	5.0	9.0	11
			17.5	$1\frac{1}{2}$x40	16.0	28.0	32
			17.5	10x600	4.0	7.0	8
			59.0	$1\frac{1}{2}$x40	3.0	4.0	5
			59.0	10x600	1.0	1.0	1.25
Silicon diffused		0.045	8.0	10x600	10.0		
		0.045	38.0	$1\frac{1}{2}$x40	20.0	22.0	24
		0.045	38.0	10x600	3.5	3.5	4.0
		0.045	72.0	10x600	1.5	1.6	2.0
		0.125	17.0	$1\frac{1}{2}$x40	170.0		
		0.125	17.0	10x600	35.0	40.0	50
		0.125	66.0	$1\frac{1}{2}$x40	45.0	50.0	65
		0.125	66.0	10x600	8.0	9.0	10

* The specimens tested were chiefly experimental or preproduction units, so do not necessarily indicate the probable performance of production units.

† Approximate diameter of effective circular junction of alloy types.

‡ Dimension of square wafer used in diffused types; approximately entire wafer area is effective junction.

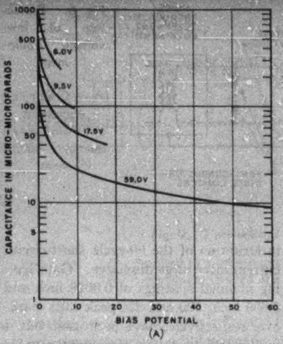


Fig. 4(A). Silicon alloy diodes, capacitances as a function of bias potential, assumed junction area = 0.0024 cm²

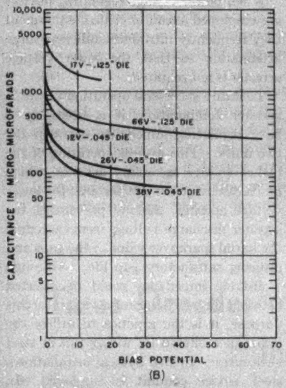


Fig. 4(B). Diffused silicon junction diodes, capacitances as a function of bias potential

mittent surges or short-duration transients, consequently, valid ratings must be determined by specific tests employing a waveshape comparable to that which the device is likely to experience in the field. For practical reasons, a unipolarity wave is generally employed in such tests and is selected on the basis of equivalent duty. Field experience has indicated that for communication applications, ratings based on a 10x600-μsec wave are reasonably conservative and satisfactory for engineering purposes. (The Western Electric Company plans to include in its published data sheets, information on the inverse-surge-current capabilities of its protection Varistors, generic term used in coding its nonlinear devices.)

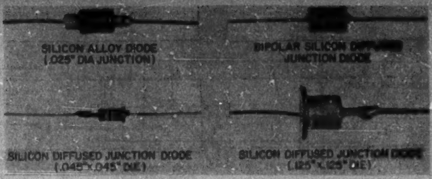

SILICON ALLOY DIODE
(.025" DIA JUNCTION)
BIPOLAR SILICON DIFFUSED JUNCTION DIODE
SILICON DIFFUSED JUNCTION DIODE
(.045" x .045" DIE)
SILICON DIFFUSED JUNCTION DIODE
(.125" x .125" DIE)

Fig. 5. Semiconductor voltage limiters scale≅full size

Test experience shows that diodes subjected to intermittent surges within their current rating do not appear to suffer any progressive degradation. If failure occurs on moderate overloads, it will be by short-circuiting, which is preferable from a protection standpoint to an open condition. Furthermore, in testing diodes in connection with establishing their duty ratings, tendency to fail will be revealed after only a few applications of a surge of critical value.

The inverse-current carrying ability of a junction is directly proportional to its area and inversely proportional to its breakdown voltage; consequently the higher voltage junctions limited in area to those presently available, 0.125 inch by 0.125 inch, have low-current ratings which restrict their application. However, diodes having lower voltage ratings may be connected in series and thereby meet higher current carrying requirements.

Table I gives the surge-current ratings of several miscellaneous silicon alloy-and silicon diffused-junction protection diodes.

DIODE CAPACITANCE

Protection diodes have capacitance which varies with impressed voltage. Capacitance variations with changes in signal level rarely present a problem in lower frequency applications or when connected at low impedance points in a circuit. In carrier applications, however, the possible effects of capacitance variation should be considered. The theory of diode capacitance is quite complex but an understanding of the general relationships is sufficient for practical purposes. In general, capacitance is directly proportional to junction area and in the inverse direction is inversely related to diode breakdown values. Also the capacitance of a given unit will decrease with the application of inverse bias voltage. These relationships are shown in Fig. 4. In the forward conducting direction, the capacitance of a junction will increase with the application of forward bias voltage. The capacitance of a no. 2052 bipolar unit at 0.0-volt bias is approximately 500 μμf and at 0.35 volt, approximately 600 μμf.

Possible adverse effects of diode capacitance on the functioning of a circuit can be more conclusively determined by test than by the theoretical approach. The development of diodes having significantly lower capacitances is receiving active attention.

BREAKDOWN-TEMPERATURE DEPENDENCE

The inverse-voltage breakdown characteristic of a silicon junction is dependent to some extent on ambient temperature, varying directly at an approximate rate of 1%/10 C (degrees centigrade) Unless otherwise stated it may generally be assumed that the breakdown rating is based on an ambient of 25 C. In protection practice this variation may be neglected since the design margins employed between the normal operating potentials of a system, the breakdown values of the protection diodes, and the withstand ability of the equipment will ordinarily be quite broad in relation to such changes.

PHYSICAL SIZE

Several types of diodes produced for surge protection purposes are shown in Fig. 5. It may be noted that these units do not employ heat sinks as in the case of diodes intended for rectifying applications since the sink would not increase their intermittent-surge-current carrying ability.

Application of Communication Protection Devices

Semiconductor diodes can be arranged to provide both longitudinal and metallic protection. They are chiefly employed, however, for metallic protection since most transmission apparatus is coupled to the line by transformers, which effectively interrupts the longitudinal path. Generally the dielectric strength of the transformers with respect to ground will exceed 600 peak volts, and discharge gaps will provide adequate voltage limitation.

Two simple arrangements utilizing the inverse breakdown characteristics of semiconductor diodes are shown in Figs.

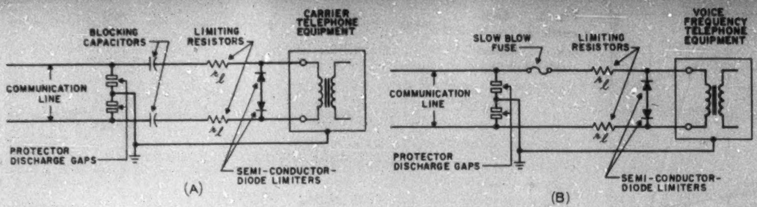


Fig. 6. Protection of low-voltage equipment

A—Carrier equipment
B—Voice-frequency equipment

6(A) and 6(B). Fig. 6(A) could be used to limit metallic voltages appearing at the line terminals of carrier equipment. The limiting resistance r_l is selected to limit surge-current magnitudes to values within the short-duration current capabilities of the diodes. The minimum value of limiting resistance that may be used depends upon diode design and the duration of the current surge. In some applications, resistance values as low as 10 to 20 ohms may be adequate. In carrier applications, a series-blocking capacitor may be used to protect the equipment against 60-cycle power currents if the exposure conditions warrant.

The second, Fig. 6(B) shows a protection arrangement for voice-frequency equipment. The only difference in this case is the substitution of a fuse for the blocking capacitor to protect the diodes against long-duration foreign currents.

In practice it will be found that considerable flexibility in diode application is possible. Since some components, such as line coils, will generally withstand surge voltages up to the operating value of the protector gaps, the diodes need not necessarily be placed at the input to the equipment but may frequently be connected in the circuit closer to the vulnerable components. In so doing, it may be possible to utilize some existing circuit impedance to limit the surge current through the diodes to a value compatible with their rating. Furthermore, it may also be possible, as shown in Fig. 7, to select a diode location having a d-c bias voltage. In this case only a single diode will be required.

An application utilizing only the forward conducting characteristic of silicon diodes for protection purposes is shown in Fig. 8. The maximum normal voice-frequency voltage is assumed to be no greater than about 0.1 volt, in which case a bipolar unit can be employed without adversely affecting system operation. Because of the anticipated low cost of such protector units, it might be economically feasible to employ several such units in series, to allow their application to a higher voltage circuit if the situation requires.

Protection Against Surges From Power Distribution Facilities

The first step in reducing the magnitude of power-line surges entering an installation is by the use of lightning arresters on the secondary service conductors. In this application the preferred location for these arresters is at the service weatherhead, for reasons to be subsequently discussed.

Standard secondary arresters are available having sparkover values in the order of 1,600 peak volts and capable of interrupting the subsequent 60-cycle short-circuit current at secondary voltages up to 175 volts rms to ground. There is a significant protection advantage in securing station service on either a 120-240-volt 3-wire or a 120-208-volt 4-wire basis. Such systems provide a grounded neutral connection for the secondary arresters whereby the potential across the arresters will not exceed their 175-volt rating. The next higher standard arrester is rated 650 volts-to-ground and requires potentials in the order of 2,800 peak volts to initiate sparkover.

Experience has shown that a 175-volt secondary lightning arrester will provide dependable protection for equipment normally considered in the power class such as motors, heaters, lamps, etc. However, such arresters do not fulfill the protection requirements for certain types of communication equipment as previously discussed. Where a high degree of operating reliability is required, it has been found necessary to employ an additional stage of protection in the form of a lower voltage arrester frequently applied to branch circuits serving critical equipment.

This low-voltage arrester is generically the same as any valve-type arrester consisting of a silicon carbide varistor, nonlinear element, in series with a standard carbon protector gap. The varistor element has a minimum resistance at 120 volts rms in the order of 15 ohms to insure interruption of the 60-cycle short-circuit current after surge discharge. Gaps having nominal spacings of 0.0028 inch and 0.006 inch, 500 and 850 peak volts sparkover respectively, have been used, but it is desirable to design equipment capable of co-ordinating with the 0.006-inch gap since it will withstand about four times greater duty than the smaller gap without permanently grounding.

When power voltage regulating devices are employed ahead of critical equipment they frequently introduce sufficient surge attenuation so that the lower voltage arrester is not required.

To insure successful operation of the 2-arrester combination, it is necessary to have regulating impedance between the two units. This insures operation of the higher voltage secondary arrester which limits discharge current through the lower voltage arrester, thereby preventing the arrester discharge voltage from exceeding the initial sparkover value of the gaps and insuring satisfactory gap life. Although a discrete impedance could be inserted between the two lightning arresters for this purpose, it is the practice to utilize the impedance offered by wiring run in steel enclosures. In Bell System installations, steel service conduit is employed. In order to secure its maximum benefit, the secondary arresters are located at the

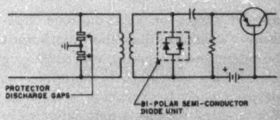


Fig. 7. Application of a single diode for protecting a transistor

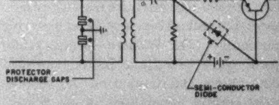


Fig. 8. Application of a bipolar diode unit for protection of transistor circuitry

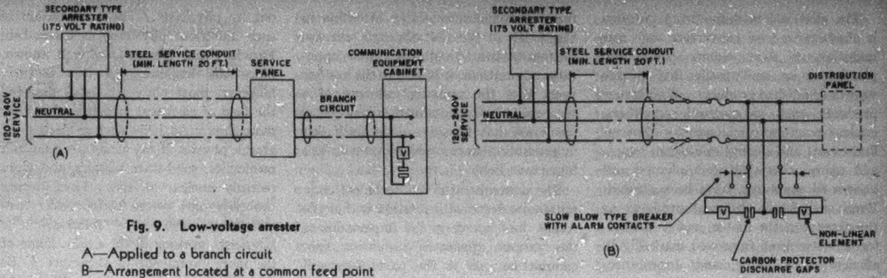

SECONDARY TYPE ARRESTER (175 VOLT RATING)
STEEL SERVICE CONDUIT (MIN. LENGTH 20 FT.)
SERVICE PANEL
COMMUNICATION EQUIPMENT CABINET
120-240V SERVICE
NEUTRAL
BRANCH CIRCUIT
V
(A)
SECONDARY TYPE ARRESTER (175 VOLT RATING)
STEEL SERVICE CONDUIT (MIN. LENGTH 20 FT.)
DISTRIBUTION PANEL
120-240V SERVICE
NEUTRAL
SLOW BLOW TYPE BREAKER WITH ALARM CONTACTS
V
NON-LINEAR ELEMENT
CARBON PROTECTOR DISCHARGE GAPS
(B)

Fig. 9. Low-voltage arrester

A—Applied to a branch circuit
B—Arrangement located at a common feed point

weatherhead. Laboratory tests and field experience show that a 20-foot conduit run will give satisfactory results. Since 20 feet is about the minimum length normally used in our installations, the low-voltage arresters may be located in the area of the service cabinets if desired. However, these low-voltage arresters are frequently placed in the equipment cabinets to secure the benefit of longer wire runs.

Two principal methods of application are used. In installations where there is spare equipment with provision for automatic switching of units, or where the rectifiers are of the trickle type and the equipment in turn operates from a battery the arresters are placed directly on branch circuits as shown in Fig. 9(A). In such cases the branch-circuit overcurrent protection device may be allowed to operate if an arrester should become permanently grounded. In situations where service continuity is not assured by the means mentioned above, the arrester circuits are provided with their own alarmed disconnect devices as shown in Fig. 9(B) to avoid de-energization of the equipment.

The use of a disconnecting device in series with the lightning arrester has the disadvantage of leaving the equipment unprotected during the time required to service the arrester, however, the situation involves a compromise and it has been assumed that the exposure of the equipment during this period constitutes less risk of service interruption than from the operation of a circuit fuse or breaker caused by a permanently grounded arrester bridged directly on the circuit.

References

1. Rectifiers in High-Voltage Power Supplies, F. W. Gutzwiller. *Electronic Design*, New York, N. Y., July 23, 1958.

2. Lightning Protection of TD-2 Stations, D. W. Bodle. *Bell Laboratories Record*, New York, N. Y., Dec. 1957, pp. 509–11.

3. Silicon P-N Junction Alloy Diodes, G. L. Pearson, B. Sawyer. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Nov. 1952, pp. 1348–51.

4. Silicon Diffused Junction "Avalanche" Diodes, H. S. Veloric, K. D. Smith. *Journal*, Electrochemical Society, Inc., New York, N. Y., vol. 104, Apr. 1957, pp. 222–26.

Three Decades of Progress in Electrical Insulation

L. J. BERBERICH
FELLOW AIEE

Synopsis: The last 30 years cover a period marked by outstanding advances in certain areas of electrical insulation. This period begins with what might be termed the synthetic-chemical age and ends with the beginning of the space age. Specifically, the paper records briefly the progress made in the theoretical phase of the field, the development of new insulating materials, the revision of the electrical insulation temperature standards, and the application of insulating materials to power equipment. The progress made in the development of materials, many of which were unheard of 30 years ago, has been truly phenomenal. This has resulted in marked advances in the design of new equipment. A few of these advances in the power field are discussed. Progress made in insulation temperature standards has resulted in new revolutionary methods of rating insulating materials and systems of materials as used in electric equipment. This has enabled the electrical industry to take better advantage of the host of new synthetic materials developed during the period covered by this review. This period may be characterized as one in which research and development in the insulation-materials area has, in a marked way, stimulated radical advances in design, reliability, and performance of electrical equipment. No radically new advances were made in the theoretical phase of the field, but the gradual progress made in the underlying theories has contributed considerably to our understanding of dielectric phenomena, especially in the liquid- and solid-state areas.

IT HAS BEEN SAID that the creation of a new product or the improvement of an existing one must be preceded by an advance in the materials which go into the making of these products. This statement is particularly true of the electrical manufacturing industry in general, which thrives on change directed toward progress. More specifically, it is particularly true of electrical insulation, since much of the progress made by the industry can be traced to advances made in the electrical insulating-materials field. For example, the motors, transformers, and other electrical components now being developed for operation at temperatures up to 500 C (degrees centigrade) for use in high-speed military aircraft, would be impossible without the recent progress made in inorganic insulating materials.

Paper 59-75, recommended by the AIEE Electrical Insulation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, February 1–6, 1959. Manuscript submitted October 20, 1958; made available for printing November 20, 1958.

L. J. Berberich is with the Westinghouse Electric International Company, East Pittsburgh, Pa.

The author wishes to acknowledge the assistance of his associates in the Westinghouse Electric Corporation. He also wishes to thank Joseph Sticher of the Detroit Edison Company for supplying some material on power cables.

The electrical manufacturing industry is characterized by converting such raw materials as steel, copper, aluminum, plastics, paper, and textiles into a large variety of finished products. Wholly new products, such as magnetic amplifiers, nuclear reactors, semiconductor devices, fluorescent and electroluminescent lamps, and many other electric appliances unknown 30 years ago, have been created. Even such basic electrical products as the generator, the motor, and the transformer have been improved markedly in efficiency, reliability, and appearance. Some of these have been so improved in performance and appearance that the present product bears little resemblance to its counterpart of 30 years ago. This dynamic mood of the electrical industry is stimulated in part by an aggressive research and development policy, particularly in the materials field, in which an important place has always been assigned to electrical insulation.

Further evidence of progress in the electrical industry is manifested in its phenomenally high rate of growth. During the last 3 decades, this growth has averaged about 10% per year. The magnitude of the growth can be better appreciated when this is compared with 3% annual increase of our gross national product. Besides the normal factors which stimulate growth, such as increase in population, there is the almost insatiable demand for electric power-consuming devices. This demand is being greatly stimulated by the continual reduction in cost, as well as improvement in reliability and performance of these devices. In all of these, the progress made in the field of electrical insulation has played an important if not commanding role.

The period covered by this review was selected for several reasons. It covers the time of the writer's research and development experience in this field; it marks the most remarkable progress in the electrical insulation area; AIEE showed increasing awareness of the importance of this field, leading to the formation of the Electrical Insulation Committee (EIC). The writer has served as chairman of this committee.

The EIC now has eight very active subcommittees. It recently organized and sponsored jointly with the National Electrical Manufacturers Association, the first National Conference on Electrical Insulation Application. It is expected that this conference will be an annual affair, and plans are underway for the 1959 conference to be held in Washington. While electrical insulation has received some considerable attention in the AIEE in the past through subcommittees associated with the various apparatus committees, it is only in the last few years that the unifying influence of a separate technical committee on electrical insulation has already made itself felt in generating more rapid progress in this important field.

The existence of a full technical committee on electrical insulation within the AIEE does not deny the importance of the various apparatus insulation subcommittees. As in the past, they still will treat important insulation problems related to the insulation systems as used in various types of electric equipment. In contrast, the EIC will treat the field of electrical insulation broadly, ranging from theory to general applications, the latter especially when more than one line of apparatus is involved. Neither does the existence of this committee in the AIEE deny the importance of the excellent work done in other technical organizations. This is especially true of the National Research Council Conference on Electrical Insulation, Committee D-9 on Insulating Materials of the American Society for Testing Materials, the Electrical Insulation Division of the Electrochemical Society, and a similar group in the American Ceramic Society. These groups treat the more basic chemical and physical aspects of the field, and supplement rather than duplicate the work of the AIEE committee and the various insulation subcommittees attached to the apparatus committees. All of these groups were founded either during or shortly before the beginning of the period under discussion.

To review in detail the progress made in this field during the last 30 years would require considerably more space than is available. Hence, this treatment of the subject will be limited to brief discussions of progress made in the theoretical phase of the field, the insulating-materials area, the insulation temperature standards area, and the application to various kinds of power equipment, such as transformers, rotating machinery, power capacitors, etc. While this review will thus be limited and also influenced by the writer's own background of experience, it is hoped nevertheless that a reasonable, though much abbreviated, picture of progress in this important area is presented.

Progress in Theoretical Area

Most of the early work on the behavior of dielectrics in electric fields was carried out by physicists. The contribution of such illustrious physicists as Cavendish, Faraday, and Maxwell are well known. Since the beginning of this century, however, most physicists have deserted the field of electrical insulation for other more glamorous fields, such as electron or atomic physics, X-ray structure, quantum mechanics, solid-state physics, and more recently nuclear physics. Thus, during the early part of the period under consideration there were relatively few physicists working in the field. Some of the more prominent of these are J. B. Whitehead (deceased), S. Whitehead (deceased), A. von Hippel, K. W. Wagner, W. Jackson, H. Fröhlich, P. Debye, L. B. Loeb, and others. Well before the beginning of this period, Maxwell had developed his theory of dielectric absorption or interfacial polarization (1875); Paschen had announced his law relating electrical breakdown of a gas and the product of pressure and electrode separation (1889); Townsend had developed his ionization-by-collision theory of breakdown in gases (1900); Debye had published his paper on his theory of polar molecules (1912); Wagner had announced his thermal breakdown theory for solids (1922); and the fields of quantum and wave mechanics were being created with the publication by Schroedinger of the wave equation (1926). No radically new theory was developed during the period, but other new developments in the fields of physics and chemistry did much to improve our understanding of dielectric phenomena, especially in the solid- and liquid-state areas.

Important contributions to our understanding of conduction, polarization, and electrical breakdown in dielectrics were made during the last 30-year period, however. While Faraday, Maxwell, and others were familiar with the fact that in insulators, in contrast to metals, electrons were not free to move except in a restricted manner, a clear explanation of the difference in conduction between metals and insulators did not develop until the advent of quantum mechanics in the early 1930's. According to this theory,[1] the valence electrons of both metals and insulators are arranged in energy-level bands, with the important difference that in insulators the conduction band is filled with electrons whereas in metals it is only partly filled. Thus, when a small voltage is applied to a metal, the energy of the valence electrons is easily raised and conduction is possible in the unfilled band. In insulators, on the other hand, electronic conduction does not easily take place because of the presence of a relatively large

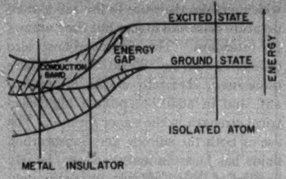


Fig. 1. Energy-band picture for metals and insulators compared with that for isolated atom

forbidden-energy band immediately above the filled conduction band. Considerable voltage, essentially the equivalent of breakdown voltage, must therefore be applied to insulators to cause electrons to bridge this region in large numbers. Most of the conduction observed in insulators at voltages well below electrical breakdown, therefore, has been observed to be ionic in nature rather than electronic. The energy-level picture illustrating schematically the difference between metals and insulators is shown in Fig. 1.

Semiconductors lie intermediate between metals and insulators and the conduction may be either electronic or ionic, depending on the nature of the semiconductor and the temperature. As a matter of fact, the whole body of knowledge developed in the field of solid-state physics which led to the development of such devices as the solid-state rectifier, the transistor, dynistor, etc., has added considerably to our understanding of conduction in insulating solids.

Knowledge of conduction and electrical breakdown in gases, in particular, has been improved through extension of Townsend's theory and study of Paschen's law[2,3] during the period. Also, it was discovered that certain gases, particularly those containing the electronegative elements of chlorine, fluorine, and bromine possess considerably higher electrical strengths than nitrogen or other similar gases. This has been ascribed to the ability of electronegative elements to capture electrons during the early stages of the breakdown process, thus making necessary the application of a high voltage to reach breakdown.[3,4,5] Sulfur hexafluoride, SF_6, and perfluorobutane, C_4F_{10}, are examples of such electronegative gases.

As a result of work done by von Hippel,[6] Fröhlich, and others, our knowledge of breakdown in solids, and to lesser extent in liquids, has been improved appreciably during the period. It appears that all breakdown in the final stages involves the formation of electron avalanches but that the processes leading to the final stages in solids and liquids differ from the simple ionization-by-collision process which applies for gases. It was also learned that breakdown in solids, in particular, is decreased by increase in temperature either through externally applied heat or internally generated heat caused by losses.[7] There are some who believe that there are two mechanisms of breakdown[8] in solids, namely: 1. electronic, which is substantially independent of temperature, and 2. thermal, which decreases linearly with temperature on a log-breakdown voltage, versus reciprocal of the absolute-temperature plot.

Much more, however, needs to be done before we completely understand breakdown in solids, and especially in liquids. In the case of liquids, in particular, many factors enter such field emission, electrode effects which handicap the investigator. Despite these difficulties, some recent careful work on hydrocarbon liquids[9] indicates that the electrical strength is a linear function of density of the liquid and also that molecular structure has an important effect on electrical strength.

Much also has been learned during this period about polarization and conduction in solids. Faraday had observed, many years ago, that when a fixed direct voltage is suddenly applied to an insulator, the current decreases with time and finally reaches a finite fixed value. It is now known that this current in the generalized case is made up of a number of components. These are as follows: 1. Electronic polarization current, which is caused by an elastic displacement of the electrons with respect to the nuclei of the atoms constituting the material. All materials show this type of polarization. 2. Ionic polarization current, which is caused by an elastic shift of atoms themselves with respect to each other and occurs especially in cases where atoms of different types form molecules in a solid. In these cases the electrons of the two or more dissimilar atoms are not shared uniformly. This effect occurs in crystals like KCl, and is particularly pronounced in a ferroelectric material such as barium titanate, whose dielectric constant is very high for this reason. 3. Orientation or dipole polarization current, which is caused by the orientation of permanent dipolar molecules in the direction of the field. An example of this is found in the cellulosic paper in which hydroxyl groups (OH) form a permanent dipole with respect to carbon atom to which they are attached and hence rotate when a field is applied. More than half of the dielectric constant of the cellulose fiber is contributed by this mechanism. 4. Interfacial or space-charge polarization current, which is caused by charge carriers or ions which can move only a limited distance or become trapped at interfaces of the various components of an inhomogeneous dielectric. Most solids show this effect in varying degrees, and under alternating voltage it produces contribution to both losses and dielectric constant. 5. Final constant conduction current, which is caused by ions migrating more or less unimpeded through the dielectric and contributes only to the losses and not to dielectric constant under alternating voltages.

All of these components go to make up the charging current when alternating voltage is applied to a capacitor in which a generalized material showing all of these mechanisms of polarization is the dielectric. The first two are generally considered instantaneous polarizations in the sense that they can respond, if present, to frequencies of 10^{12} cps (cycles per second) or higher. The next two seldom are observed at frequencies above 10^8 cps, and in many cases do not respond at much lower frequencies. The last, of course, is not a polarization current, therefore, does not contribute to the dielectric constant but produces losses over the whole frequency range.

It should be pointed out that during the period a considerable number of physical chemists have worked in this field. Among the better known might be mentioned Kirkwood, Debye, Symth, Onsager, and Fuoss. These and many others have made very valuable contributions in relating chemical structure to dielectric properties, especially in respect to dielectric constant and loss behavior. This type of information has been valuable in the development of new dielectrics, particularly for capacitor, cable, and other applications.

Engineers were also engaged in this area of research. They were especially concerned with refining the various theories developed by the physicists and physical chemists and applying their findings to the solution of practical problems. As was pointed out in 1940,[10] and it is equally true now, progress in this field is due to the combined efforts of physicists, chemists, and engineers, the "triple alliance in electrical insulation."

This brief review of the theoretical aspects of the field may be summed up by saying that no radically new theory has been developed during the period and that steady progress has been made in improving our understanding of dielectric phenomena. It is also pointed out that our knowledge in the insulation area has been considerably enhanced through advances in related fields, particularly in solid-state physics and chemistry. Marked contributions were also made through relation of structure to dielectric properties.[11] Many challenging problems remain to be solved especially in the electrical strength and conduction areas. A more extensive review of this phase of the subject is given by von Hippel in reference 3.

Progress in the Materials Area

The progresss made in creating new materials during this period has been truly remarkable. The fact that the period began with what might be called the synthetic age has already been mentioned. It is in this area that organic chemists, working not only in chemical industry but also in the electrical industry, have played a stellar role. In the early 1920's organic chemists were almost unheard of in electrical research and development laboratories. They, however, began to appear in increasing numbers in the early 1930's. The result of their work, and particularly of those in the chemical industry, has been an almost continual stream of new materials which might be called the products of synthetic chemistry. They have largely replaced the natural materials in common use in the early days of the electrical industry. Even though in some cases new synthetic materials with the desired electrical properties were developed in the electrical industry, the chemical industry must be credited with the major contribution, particularly in the plastics and elastomers areas.

In order to obtain a better picture of the progress made over the last 30 years, it may be of interest to examine lists of the more important categories of materials available in 1928 as compared with those available now. The following tabulations contain such lists in much abbreviated form.

Gaseous Materials. Table I shows that the most important new addition to the list of gaseous materials in the last 30 years center around fluorine containing compounds known as electronegative gases. Only several of the available gases of this type are listed. The principal advantage of these over ordinary gases is higher electrical strength by a factor of two or more at the same pressure.[3,4,5,12] Increasing interest is being developed in these gases for uses in special gas-filled transformers and to some extent in capacitors. Sulfur hexafluoride more recently has been found to possess the additional advantage of high arc-interrupting ability compared with air and other gases. This combined with high electrical strength has resulted in its use in a new line of circuit breakers.[12] The use of hydrogen in 3,600-rpm turbine-generators as a cooling and insulating medium was initiated at the beginning of the period, and it continues to be used, at higher pressures and with greater effectiveness. The principal advantages of hydrogen in this application are better heat transfer and lower windage losses. Its electrical strength unfortunately is about 60% that of air at the same pressure.[13]

Table I. Progress on Gaseous Materials

1928	1958
Air	Air
H_2	N_2
CO_2	CO_2
H_2	H_2
	SF_6*
	C_2F_6, C_4F_8, . . . †

* Sulfur hexafluorides.

† Fluorocarbons.

Table II. Progress in Fluid Materials

1928	1958
Rosin oil	Petroleum oil
Castor oil	Chlorinated hydrocarbon fluids
Petroleum oil	Diester fluids
	Silicone fluids
	Fluorocarbon fluids

Fluid Materials. Table II shows that petroleum oil is the only important fluid in use at the beginning of the period that is still in use now. This low cost fluid, which has been much improved during the period,[3,14] still is the most common fluid used in high-voltage transformers, circuit breakers, and oil-impregnated paper cables, and to a small extent in capacitors. Rosin oil and castor oil are no longer being used to any appreciable extent and have been replaced by the synthetic fluids, especially in capacitors. The chlorinated fluids, known by the generic name of "askarel" and developed in the early 1930's,[15] possess fire-retardant properties and have more than twice the dielectric constant of petroleum. This has enabled them to find extensive use in power capacitors and to a limited extent in transformers requiring a fire-retardant fluid. The diester fluids have been useful in high-frequency (10^5 and 10^6 cps) capacitors.

The silicone fluids,[3] developed during World War II, possess thermal stability which enables their use to the 180 C temperature class and higher but they are not fire-retardant. The fluorocarbons, also developed during World War II, are quite inert, thermally stable, fire-retardant, and in addition, possess high electrical strength in the vapor form.[3,4] The use of both the silicone and fluorocarbon fluids has been limited, however, largely because of their high cost, approximately 100 and 300 times the cost of an equal volume of petroleum oil, respectively. It should be pointed out, however, that part of the reason for the higher cost, particularly in the case of the fluorocarbon fluids, is because they are still only in pilot production. The askarels first introduced over 25 years ago are made in large quantities and sell at from five to eight times cost of an equal volume of petroleum oil.

Papers and Textiles. Table III shows that a considerable number of additions have been made during the period to the list of the familar cellulosic and asbestos papers and fabrics. Cellulose, however, in the form of paper, pressboard, cotton fibers and fabrics, is still most important in this category from a volume standpoint. The materials based on cellulose combine low cost with excellent electrical, chemical, and physical properties. They, however, have the disadvantages of being sensitive to moisture and of being limited thermally to the 105 C temperature class.

It is this thermal limitation, as well as sensitivity to moisture, that stimulated the development of the inorganic-fiber papers and more recently organic-synthetic-fiber papers and fabrics. The former are satisfactory for use to at least the 180 C temperature class. Usually, however, their temperature rating is determined by the resinous binder used with them. The more important examples of these are the glass-fiber fabrics, mica paper, and glass-fiber paper. The synthetic organic papers are still in the pilot-production stages although most of the same fibers, such as nylon (polyamide), Dacron (polyethylene terephthalate), and Orlon (polyacrylonitrile) have been produced in large volume in the form

Table III. Progress in Papers and Textiles

1928	1958
Cellulosic papers	Cellulosic papers
Cotton fabrics	Cotton fabrics
Vulcanized fiber	Vulcanized fiber
Asbestos papers and fabrics	Asbestos papers and fabrics
	Synthetic papers and fabrics
Silk fibers and fabrics	Inorganic fibers and fabrics
	Inorganic papers

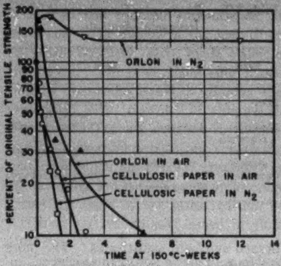


Fig. 2. Effect of aging at 150 C on cellulosic paper and synthetic acrylonitrile (Orlon)

of fibers for use in woven fabrics. The high cost and the lack of voltage-barrier action in these and all other fabrics have, however, seriously limited their use as insulating materials. The chemical and paper industries are now learning how to make papers from the synthetic fibers which promise to have at least some voltage-barrier action and considerably lower cost than their woven counterparts. This should result in numerous applications when they become available in production quantities at a reasonably low price. The thermal-stability data of Fig. 2 comparing Orlon fibers with cellulosic paper show that the former should have considerable advantage over the latter, especially if sealed away from air.

Resins and Varnishes. This is an area where synthetic materials have made very considerable inroads on the formerly used natural materials, such as rosin, natural gums, shellac, and asphalt as shown in Table IV. This class of material has found extensive use as adhesives, binders, varnishes, impregnants, and casting molding and laminating resins. Some are also used in the form of extrusions, especially on wires and cables.

Table IV. Progress in Resins and Varnishes

1928	1958	
Rosin	Shellac	Polyester resins
Natural gums	Asphalt	Polyvinyl chloride
Coal tar pitches	Vegetable drying oils	Epoxy resins
Shellac	Phenolic resins	Methyl methacrylate
Asphalt	Alkyd resins	Melamine
Vegetable drying oils	Polystyrene	Polyamide resins
Phenolic resins	Polyethylene	Polyurethane resins
	Polyethylene terephthalate	Silicone resins
	Cellulosic resins	Fluorocarbon resins

Of particular importance is the development of a number of resins, specifically the polyesters, epoxies, and certain silicones, which are fluid in the unreacted form, thus requiring no solvents. These are completely heat-reactive and do not evolve volatile by-products while reacting to the solid state. This enables their use in many casting and impregnating applications where obtaining void-free structures is important. No resins of this type were available as early as 1935 and all were developed within the last 20 years.

Perhaps of even greater importance is that some of these resins have enabled operation of electrical equipment at higher temperatures than the original 105 C and 130 C classes, classes *A* and *B*, respectively. As a matter of fact, the present AIEE 180 C temperature class, class *H*, was originally designed around the silicone materials which were developed during World War II. The fluorocarbon resins, also developed during World War II, meet the requirements of this temperature class from a thermal-stability standpoint, but have not been applied as widely as the silicones in electrical equipment because of higher cost and difficult processing. However, where outstanding inertness and solvent resistance are special requirements the fluorocarbons have no equal.

The organic chemist has been expending considerable effort in the area of tailoring resin molecules to meet special requirements. Thus, the industry now has available fire-retardant resins (polyvinyl chloride), arc-resistant resins (melamine, methacrylates, some alkyds, and ureas), and thermally stable resins (silicones and fluorocarbons). It is in this area that molecular engineering, or designing molecules for particular applications, is meeting with considerable success and more is expected in the future.

Wire Enamels. The wire-enamel field is one in which the chemist practices molecular engineering to a high degree. The thin film placed on copper or aluminum wire must meet extemely rigid requirements of flexibility, toughness, solvent resistance, and electrical strength. Table V shows that in 1928 only one type of wire enamel made from natural resins was available. The first synthetic wire enamel was developed in the late 1930's by the General Electric Company.[16] It was found that, on reacting a polyvinyl formal resin developed by Shawinigan Chemical Company with a phenolic resin, an enamel with considerably better properties than those of the older oleoresinous enamels resulted. Other synthetic enamels soon followed, and in the last 10 years a considerable number of new ones were developed with even better properties, especially in respect to thermal stability. The curves of Fig. 3, giving the lives of the various types of enamels as a function of temperature, show the progress that has been made in respect to thermal stability. These curves were developed by an AIEE working group on enameled magnet wire,[16] which currently is very active in developing test procedures for determining the thermal rating of film-coated magnet wire. The data of Fig. 3 are for enamels with no varnish treatment. Some of these enamels show even greater lives when treated with compatible varnishes, and data on varnished enameled magnet wire are also included in the report of the AIEE Working Group.[17]

Table V. Progress in Wire Enamels

1928	1958	
Oleoresinous	Oleoresinous	Polyesteramide
	Polyamide	Polyester-epoxy
	Polyvinylformal	Polyester
	Epoxy	Modified silicone
		Polyurethane

Elastomeric Materials. This is another area where the organic chemist has made remarkable progress. Table VI shows that only several materials, namely natural rubber and gutta percha, were available in 1928 and that all of the various synthetic rubbers have since been developed.[8] Work was being done on synthetic rubbers in Germany in the 1920's, but not much progress was made in this country until the late 1930's when it became apparent that our source of natural rubber supply might be jeopardized by the impending World War II.

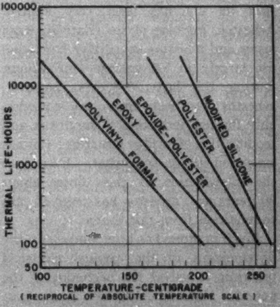


Fig. 3. AIEE No. 57 twist-test results on various classes of unvarnished wire enamels

Table VI. Progress in Elastomeric Materials

1928	1958, Rubber	
Natural rubber	Natural.......	Butyl
Gutta percha	Neoprene.....	Polyester
	Buna-N.......	Silicone
	GR-S..........	Fluorocarbon

The exception of this was neoprene which was first synthesized at Notre Dame University in the early 1930's and then developed commercially by Du Pont.

Some of the synthetic rubbers listed were developed in Germany but a good proportion of them, particularly butyl, silicone, polyester, and fluorocarbon rubbers were developed in this country. While most of the elastomers listed are limited to the 105 C temperature class, the silicone rubbers in particular have been found to meet the requirements of the 180 C temperature class. Perhaps the single most important application of synthetic rubbers in the electrical-insulation area is the insulated wire and cable field.

Ceramic and Other Inorganics. Table VII lists only the major classes of materials of this type. Again it is seen that a considerable number of wholly new materials have been developed in the last 30 years. The ceramist and the inorganic chemist working with engineers have been responsible for progress in this area.[3,18] Mica, asbestos, electrical porcelain, certain electrical glasses, and cements have been used almost from the beginning of the electrical industry. Mica, in particular, has such a unique combination of properties that it has been found to be almost indispensible for the insulation of commutators and coils for the larger sizes of rotating machines. Only in recent years has mica been produced synthetically, and then only with such small flake size that its use has been very limited. Asbestos, because of its relatively high losses and moisture sensitivity has found use largely in low-voltage applications. During the last 20 years, however, asbestos has been improved considerably through reduction of impurities and modification of its fiber structure; this has expanded its use somewhat. Ever since the introduction of glass fibers about 20 years ago, asbestos has received stiff competition from this source.

Much progress has been made since the development of the original electrical porcelain in use at the beginning of the period. Not only has this early porcelain been improved but wholly new compositions, such as zircon and alumina porcelains, have been developed.[3] Too frequently it was not realized that the early electrical porcelains were not very satisfactory at temperatures much above 100 C because of their high conductivity brought about by a high-alkali-metal content. The zircon and alumina porcelains were developed for higher temperature applications and in general have low-alkali-metal contents. The latter, for example, has found extensive use in aircraft engine spark plugs, where mica used to be the only satisfactory material. Zircon porcelains, as is the case with electrical steatite, possess low losses and hence have found use in high-frequency applications.

Many new glass compositions have also been developed during the period. Those with low losses and the desired physical properties have made possible the development of the vacuum tube art as we know it today. In this connection the Corning Glass Company has played a leading role.[19]

Perhaps the most startling development in this area concerns the titanate ceramics. It has been known for many years that TiO_2 in the rutile crystalline form possesses a dielectric constant of the order of 100.[20] It was not until the early 1940's, however, that it was discovered by Wainer and associates,[21] then with the Titanium Alloy Manufacturing Division of National Lead Company, that certain titanates possess even higher dielectric constants. Values of dielectric constants of over 5,000 have been reported for barium titanate and combinations of barium and strontium titanates. Commercial compositions which have found applications in the electronic capacitor field usually consist of barium titanate with minor additions of other materials to flatten out the sharp peak in the dielectric constant versus temperature curve. The titanates also have found application in the ultrasonic field as transducers after it was discovered that they could be polarized to produce piezoelectric properties.[3]

This brief review of new materials developed during the period may be summarized by saying that outstanding progress has been made. A host of synthetic materials have appeared which have replaced many of the natural materials once used in the electrical insulation area.

Many of these materials have made possible remarkable advances in equipment design, some of which will be illustrated in the last section of this paper. A few natural materials, however, are still being extensively used because of their unique combination of properties and difficulty of economically creating improved synthetic substitutes. The more important of these are mica, petroleum oil, and cellulosic paper. Only small progress has been made in developing improved synthetic replacements for materials of this type but work now going on in the development of synthetic-organic-fiber papers appears very encouraging.

Table VII. Progress in Ceramics and Other Inorganic Materials

1928	1958	
Mica	Mica..............	Aluminum porcelain
Electrical porcelain	Asbestos..........	Titanate ceramics
lectrical glasses	Electrical porcelain.........	Electrical glasses
Cements	Low-loss steatite..	Glass bonded mica
	Zircon porcelain...	Cements

Progress in Insulation Temperature Standards Area

The problem of rating electrical equipment, and the influence of the thermal stability of materials on the rating, has been under periodic review for many years. AIEE No. *1* was the first formal document published by the AIEE and it may be best described as a guide to the preparation of insulating material and equipment temperature rating standards. Even though it was revised several times during the period under discussion no very radical changes from the 1930 edition were made until the appearance of the 1957 edition.[22] The revolutionary changes made in this latest edition were brought about by the appearance of the multitude of new synthetic materials discussed in the previous section. These newer materials could not be properly classified thermally under the old system.[23]

Under the former system, materials were placed into various temperature classes, according to whether they were organic or inorganic, or a combination of these two broad chemical classifications. Since the silicones did not fit into either of these two classifications, a new temperature class, 180 C, was added in 1947 when these important materials were finding increasing number of applications in the insulation field. A real dilemma, however, was created when it was found that an increasing number of the new synthetic organic materials possessed thermal stability comparable to that expected of combinations of inorganic and organic materials. More recently, some have even been developed which approach the silicones in this respect. This made it almost mandatory to develop a new ap-

Vol.
Pt. 1
1959

proach to the thermal classification of materials. The events of the last 10 or 15 years, therefore, have made obsolete the former system of classifying materials by definitions based on whether the materials were organic or inorganic. This turn of events resulted in a completely new philosophy of classifying materials which has evolved gradually in the last 20 years.[24] This new philosophy forms the basis of the latest edition of AIEE No. *1*, which was issued in June 1957.[23]

The beginning of the new philosophy of insulation classification dates back to the period, when Montsinger[25] and others tried to relate the effect of time and temperature to the properties, or more particularly to the life, of insulating materials. This soon led to the well-known "8-degree rule" for oil-impregnated transformer insulation, which states that the rate of deterioration is doubled for every 8 C rise of temperature. Expressed in another manner, this means that the life of transformer insulation is halved for every 8 degree rise in temperature. This important finding, though only approximately true, was soon applied to the loading of transformers when it was recognized that a short-time overload would not destroy the transformer, but merely reduce the normal life by an amount which could be computed from the temperature and time duration of the overload. The procedure was later extended to other apparatus although not covered by formal standards.

This rule-of-thumb relationship between life of insulation and temperature was used in varying ways until the late 1940's. It was then found, as the chemist had long since known, that the chemical-reaction-rate theory as exemplified by the classic Arrhenius equation may be applied to insulation aging. If it is postulated that insulation deterioration approximates a first order chemical-reaction rate, as has been shown to be true in many cases, the Arrhenius equation may be expressed as follows:

$$\text{Deterioration rate} = A\epsilon^{-B/T} \qquad (1)$$

or, since life L is reciprocally related to the rate of deterioration, life may be expressed as

$$\text{Life} = L = A'\epsilon^{B/T} \qquad (2)$$

which may be converted to the logarithmic form, or

$$\text{Log } L = \text{Log } A' + B/T \qquad (3)$$

where T is the absolute temperature, A, A' and B are constants determined by the activation energy of the particular reaction, the gas constant, and the concentration of the reacting materials. Thus, equation 3 indicates that if the logarithm of the life of insulation is plotted against the reciprocal of the absolute temperature, a straight line should result.

This relationship was confirmed, with few exceptions, by data obtained on many materials. It was found, for example, that when the logarithm of the life of chlorinated diphenyl-impregnated paper capacitors is plotted against the reciprocal of the absolute temperature, almost perfect straight lines resulted.[26] The exceptions were found when the second and higher order reaction rates enter into aging phenomenon which sometime take place at the higher temperatures. Extensive application of reaction-rate theory to insulation deterioration was not made until relatively recently. This was further stimulated by two important general papers on this subject, the first by Dakin[27] which appeared in this country in 1948, and the second by Mamlow[28] which appeared in Sweden in 1949. These two papers thoroughly review the principles upon which the Arrhenius equation is based and show how they apply to the analysis of insulation-deterioration data.

In the last few years these principles have been incorporated into accelerated life tests which have formed the basis of procedures for determining the thermal rating of materials[29,30] and systems of materials as they are used in various types of electric equipment.[31–35] These procedures all require the obtaining of aging data under prescribed conditions and plotting the statistically analyzed data on a log life versus reciprocal of the absolute-temperature plot. Then, if it is established that the data form approximate straight lines on this type of plot, as often is the case, the line may be extrapolated with caution to a generally accepted standard life value and the corresponding temperature rating may be read from the plot. Thus, the approximate rule-of-thumb relations, having no scientific basis, have now been almost entirely replaced by the more accurate relation based on the classic Arrhenius equation. It has also been realized only recently that the temperature-rise interval over which the insulation life is halved is not a constant, but increases with the temperature for a given material, provided a single reaction-rate law is obeyed over the whole temperature range.

The very considerable interest in this field is not limited to this country, but a similar interest has been generated on a world-wide basis by the International Electrotechnical Commission (IEC). This world standardizing body has also engaged in a review of its insulation temperature standards.[36] The wording of the definitions of various insulation temperature classes in IEC Publication 85 are very similar to those in AIEE No. *1*, illustrating the important role played in world standards by the AIEE.

The revised insulation-temperature standards now permit the electrical industry to take better advantage of the many new synthetic materials with improved thermal stability, such as the wire enamels in Fig. 3. To implement the new philosophy of insulation classification, a test procedure for each major class of material will be required, as well as a test procedure for each of the various insulation systems as they are used in electrical equipment. The work of developing these test procedures is well started, but much more remains to be done to complete this monumental task. The progress made so far, however, can be recorded as a major advance in this very important area of the electrical field. One of the more important by-products of this work is marked advance in our understanding of the mechanism of thermal deterioration of various types of materials.

Applications to Power Equipment

Reference was made earlier to the fact that much of the progress made in electrical apparatus can be attributed to advances made in electrical insulation. It will be the purpose in this section to cite a few examples of such progress in the rotating-machinery, transformer power-capacitor, and cable fields.

Rotating Machinery. In the case of the smaller 105 C class *A* machines, the more important insulating materials at the beginning of the period under review were cotton as conductor insulation, paper as slot liners or coil wrappers, and varnished cambric or untreated cotton tapes as coil insulation. The whole motor was treated in a oleoresinous or asphaltic type of varnish. With the advent of synthetic materials in the last 10 years or more, most of the cotton-covered wire has been replaced with enameled wire and part of paper slot liners and coil insulation has been replaced with synthetic films, such as Mylar (polyethylene terephthalate). The whole machine is treated with a synthetic varnish and then baked at an elevated temperature. Fig. 4 shows the improvement in motor life made possible by improvements in wire enamels and varnishes. The data were obtained from motor tests conducted according to AIEE

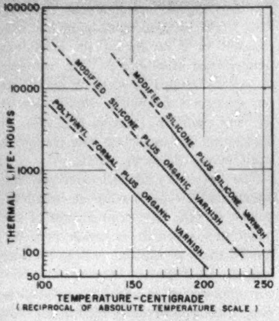


Fig. 4. AIEE No. 510 motor-life-test results on various wire enamel-varnish combinations

No. *510* random-wound motor-insulation test procedure. It is apparent from Fig. 4 that the newer modified silicone enameled wire-varnish combinations should be satisfactory for operation at 130 C class *B*, and possibly higher temperatures. The polyvinyl formal-organic varnish combination has been used since about 1940 as 105 C temperature class insulation and serves as a basis for comparison.

Mica tapes and wrappers have been used for many years in the larger motors designed for 130 C class *B* operation. Mica, however, had to be bonded with organic varnishes and this limited the use of this combination to the 130 C temperature class. With the advent of silicone varnishes, glass-fiber tapes, and glass-fiber-covered conductors, a new insulation-temperature class was created, namely the 180 C class, class *H*. Fig. 5 shows the size reduction possible in 7.5-hp (horsepower) 3-phase motor by the application of the 180 C class materials over the normal 105 C class materials. All of the 180 C class materials were developed during the period under review. More recently, some organic materials have been developed with improved thermal stability and this appears to make possible their use at temperatures above the 130 C class. For this reason, a new temperature class, namely the 155 C class, class F, was added to the 1957 edition of AIEE No. *1*. This permits the use of materials at 155 C maximum hottest-spot temperature, or higher if they meet the requirements of this temperature class by standard test or performance.

Improvements in the large power-generating units, particularly the steam-turbine-driven generators, have been even more startling. This has manifested itself in the ever increasing kva rating of the units being built. Fig. 6 shows that the maximum available kva rating in 1928 was less than 50,000 and now units approaching 500,000 kva are being built. It is true that much of the progress in this area is the result of improved cooling efficiency, but improvements in insulation also played a very important role. The use of hydrogen as cooling gas started at the beginning of the period under discussion, and marked improvements in cooling efficiency have been made by gradually raising the pressure, and more recently, by going to the inner-cooled design. In the latter, hydrogen is circulated through tubes in close thermal contact with the copper conductors of the coils, thus very effectively shortening the path for the heat produced by I^2R (current resistance) losses in the copper.

The insulation used in the stator coils of both the large turbine-generator units as well as water-wheel-generator units has been based on mica since almost the beginning of the electric power industry. The first binder used with mica flakes was shellac, but at the beginning of the period under discussion petroleum asphalts came into the picture. The latter resulted in a more flexible insulation which could be applied continuously in tape form.

In the last 15 years, however, it became apparent that as the machines became larger, the asphalt binder has some limitations, both electrically and mechanically. Since asphalt is very viscous, even when heated, it is difficult to impregnate into the coil structure to fill all of the void spaces. Also, since asphalt may be looked upon as a viscous liquid even at ordinary temperatures, this material allowed insulation movement during operation of the machine as a result of temperature cycles. It became apparent as early as 1940 that a better material than asphalt for use as a mica bond must be found and therefore active work was carried on with the increasing number of synthetic resins appearing even at that time.

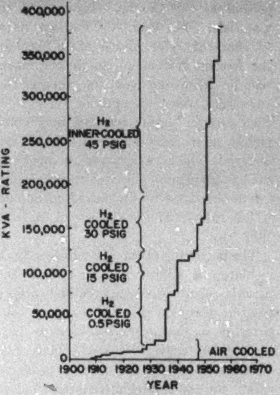


Fig. 6. Maximum available kva rating of 3,600-rpm turbine-generators

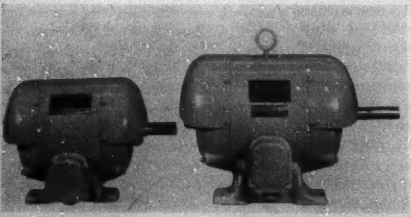


Fig. 5. Size comparison of a 7.5-hp 440-volt, 3-phase class H motor (left), and a class A motor of the same rating

After some 10 years of work with synthetic-resin-impregnating media and the development of radically different processing methods that were required, the first new synthetic resin bonded and impregnated mica insulation for large turbine-generators was announced in 1950 by Westinghouse Electric Corporation. The account of this development was published in the 1951 AIEE *Transactions*.[37] This insulation is based on the use of completely heat reactive synthetic resins which are applied in the fluid state and subsequently polymerized after they are in place. This insulation, called Thermalastic, has since been applied to water-wheel generators as well as to motors and generators in the integral hp class. Improved electrical and physical properties have enabled this insulation to achieve an outstanding record in the 8 years since its introduction. But more important, it has enabled the construction of larger and more efficient generating units with added assurance that insulation will perform more satisfactorily. The improved voltage endurance behavior of the new insulation compared with the old asphalt bonded type is illustrated in Fig. 7. A photograph of one of the first turbine-generators

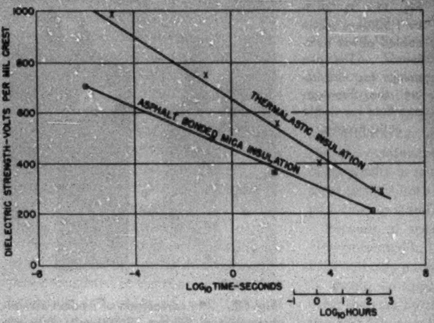


Fig. 7 (above). Dielectic strength and voltage-endurance behavior of the old and new generator-insulation systems

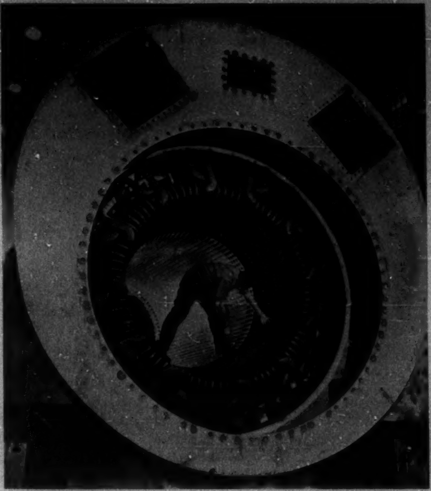


Fig. 8 (right). Hydrogen-cooled turbine-generator showing Thermalastic coils in place

constructed with Thermalastic insulation is shown in Fig. 8.

Progress is continually being made in rotating-machinery insulation and periodic announcements of new insulation systems have been made, particularly in the last few years. Among the more important of these are an all-silicone rubber insulation for motors and generators known as Silco-Flex, announced by Allis-Chalmers in 1955,[38] and a synthetic-resin bonded-mica paper insulation for large generators announced by the General Electric Company in 1958. The properties of the latter insulation, referred to as Micapal, has been described in reference 39. These few examples should serve to illustrate that insulation developments played an important role in bringing about progress in the rotating-machinery field. Some of this progress is reviewed in several relatively recent books on this subject.[40,41]

Transformers. As with rotating machinery, the progress made in transformers was to a marked degree influenced by insulation developments. In the distribution transformer area the use of enameled wire in place of cotton- or paper-covered wire, together with improvements in oil and magnetic steel, has resulted in a significant size reduction for a given rating. Fig. 9 shows photographs of one of the first type CSP 5-kva transformers, built in 1935, compared with similar 15-kva units built recently. In the lower voltage rated instrument transformers there is a trend away from the taped and asphalt-impregnated structures to resin embedded and molded units. Filled epoxy resins are finding application in cast units; there is considerable interest in the use of butyl, polyester, and silicone rubbers in molded units. Since some of these compounds have excellent weather resistance, it has been possible to replace the earlier asphalt-impregnated units with molded or cast units even for outdoor applications. In some of these the metal case and porcelain bushings were eliminated. The use of butyl rubber in this type of application was first announced in 1955.[42] A polyester rubber developed within the last few years has found considerable application in molded indoor-current transformers. Fig. 10 shows a series of the new polyester-rubber molded units, with the older units which they replaced immediately behind them.[43]

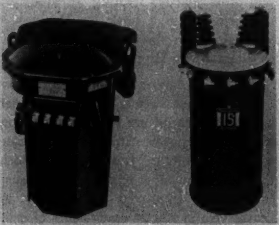


Fig. 9. Comparison of a 1935 model, 5-kva CSP-type transformer (left), and a 1958 model, 15-kva CSP-type transformer

The advent of silicones has enabled the design of smaller, more compact dry-type power and network transformers. In these units the more important materials used are fiber-glass conductor insulation, ceramic-coil spacers, and glass-mica barrier tubes. The whole unit is impregnated in silicone varnish. Prior to the introduction of silicones, no dry-type units for operation above approximately 150 C were available. Now temperature ratings have been increased to 200 C and higher.

In liquid-filled transformers, progress of a different type has been made. As in 1928, the most important insulating materials used in this type of transformer are still petroleum oil and two cellulosic products, namely, paper and pressboard. Despite considerable effort, no economically feasible synthetic equivalents of these materials have been developed. Improvements, however, have been made in both the oil and cellulosic products. These, combined with improvements in magnetic steel and cooling efficiency, have enabled the construction of units rated up to 400,000 kva. These improvements have brought about very marked reduc-

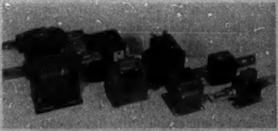


Fig. 10. New polyester rubber molded current transformers (front row), with their older counterparts

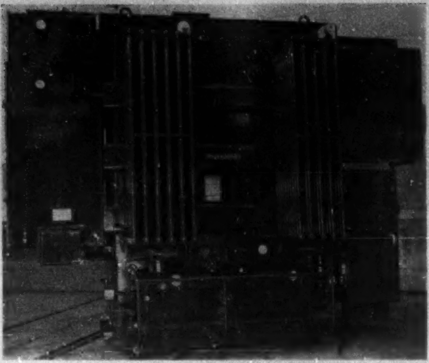


Fig. 11. A 34.5-kv 7,500-kva vapor-cooled power transformer. The two pumps for circulating the fluorocarbon fluid are visible at the bottom

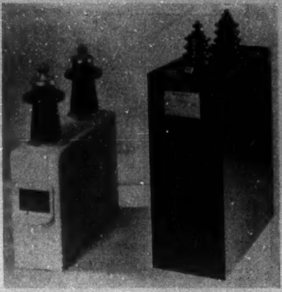


Fig. 12. Size comparison of modern askarel-impregnated 10-kva 2,400-volt 60-cycle capacitor (left), and old oil-impregnated capacitor of the same rating

tion in weight per kva as well as physical size for a given rating. This latter is important, particularly since the maximum rating that can be built is not determined by the manufacturing limitations, but by the facility with which the unit can be transported to the site of application.

With the advent of the electronegative gases in the 1940's, considerable thought has been given to applying them to transformers. With the higher electrical strength values obtainable with these gases than with air or nitrogen at the same pressure, it appeared that transformers should be an ideal application. While a specialized application of sulfur hexafluoride, SF_6, in high-voltage X-ray transformers was announced by General Electric in the 1940's, further large scale application of electronegative gases in transformers has been slow. This perhaps is because of the many problems involved in their use, not the least of which is cooling. With the development of the fluorocarbon gases and liquids during the latter part of World War II,[3,4] however, a new concept of transformer cooling and insulating was developed. This involved the spraying of an inert, fire-retardant fluorocarbon fluid on the core and coil structure, thus utilizing the heat of vaporization in cooling the transformer and at the same time making use of the high electrical strength of the electronegative fluorocarbon vapor.[4,5] The first experimental model of this transformer was constructed in 1951 and announcement of the vaporization-cooled design was made by Westinghouse at that time.[44] Several commercial units have been built, the first of which rated at 15 kv and 500 kva, has been in successful operation for more than a year. The second unit, rated at 34.5 kv and 7,500 kva, is shown in Fig. 11. Others are in the design stages.

Interest in gas- and vaporization-cooled transformers continues at a high level. A number of manufacturers are developing new design information. Several units rated at 2,000 kva and 69 kv based on the use of sulfur hexafluoride gas are under development by the General Electric Company.[45] Other gas- and vapor-cooled units of even higher rating are certain to follow. This is a striking example of how the development of new insulating materials has made possible radically new equipment design.

Power Capacitors. At the beginning of the period under review, power capacitors were constructed with cellulosic paper impregnated with a petroleum oil as the dielectric. During the period, as mentioned previously, the askarel liquids[15] which consist of chlorinated aromatic hydrocarbons, were made available. These materials, developed and produced by the Monsanto Chemical Company and its predecessors, have excellent electrical properties for this application. Of particular importance is the fact that their dielectric constant is about twice that of petroleum oil. These liquids first came into use in the early 1930's and gradually replaced petroleum oil as a power-capacitor impregnant almost completely by 1940.[15] By virtue of the fact that the dielectric constant is higher and closely approaches that of the cellulose fiber in paper than is the case for petroleum oil, the askarels have made possible a marked size reduction. This is shown in Fig. 12, which gives a size comparison between an old oil-impregnated unit and a modern askarel-impregnated unit of the same rating.

While cellulosic paper is still used in power capacitors, considerable improvements in electrical strength and dielectric losses have been made in the last 30 years. This, combined with use of askarel instead of oil, has made a possible operation of the capacitor dielectric at stresses up to 400 volts per mil, 60-cycle alternating current, and over 1,000 volts per mil direct current. In addition, lower losses have enabled the construction and satisfactory operation of units rated at 100 kva. This should be compared with a maximum rating of about 15 kva available in 1928.

The application of askarel to capacitors is a striking example where the use of a synthetic material with a cost of five or six times that of the petroleum oil it replaced, has resulted in lower over-all cost of the finished unit. Since the volume of the capacitor for a given rating has been reduced to one half that of the older, oil-impregnated unit, not only is half as much liquid required but also only one half as much paper, aluminum foil, steel for the container, etc. The net result is a more reliable unit which can be marketed at an economic price. These and other matters relating to capacitors are reviewed in a book by Marbury.[46] Capacitors are also discussed in several chapters in von Hippel's relatively recent book.[3]

High-Voltage Cables. Cellulosic paper and petroleum oil are still being used in high-voltage power cables as is the case with high-voltage transformers. Both the oil and the paper have, however, been considerably improved in the last 30 years. No synthetic equivalents of cellulose paper and petroleum oil have yet been found which combined low cost and the electrical properties required in high-

Vol.
Pt. 1
1959

voltage power cables. Plastics, such as polyethylene and butyl rubber, however, have made inroads in the medium-voltage power cables rated up to about 15 kv. Experimental installations of polyethylene-insulated power cables rated at 34 kv have been made recently. Polyvinyl chloride and rubbers of various types have been used for some time in low-voltage cables and insulated conductors. All of the plastic materials and rubbers, except for natural rubber, were developed during the period under review.

The most important breakthrough in the high-voltage cable area during the period consisted of the control of internal ionization by means of new types of cable designs. Just a little over 30 years ago, the so-called oil-filled Pirelli-type cable was developed in Italy, which enabled operation of cables up to 132 kv as early as 1927. In contrast the older solid-type cable could not be operated satisfactorily above about 69 kv. This was accomplished by use of oil channels in the cable and associated expansible oil reservoirs together with a freely flowing oil at all cable-operating temperatures.

Later, a number of other new cable designs were developed such as the self-contained compression types, the pipe-types in which the cable is kept under pressure by use of either gas or a liquid, and the gas-filled types. This has resulted in cables which are now operating in this country at voltages up to 230 kv. An experimental field installation is planned for 1959 at Cornell University to study cable designs for eventual use in the 300- to 400-kv range.

In summary, progress in cables for high-voltage power applications is not characterized by the use of radically new materials. Significant improvements, however, have been made in both the paper and oil properties and even more so in cable designs. Plastics and synthetic rubbers have almost entirely replaced natural rubber and varnished cambric formerly used in medium- and low-voltage cables. Progress in this, as with the other fields discussed in this paper, are recorded in the AIEE and other technical literature. Some of the problems encountered in the cable field have been reviewed in von Hippel's recent book.[3]

Summary

The outstanding progress made in the field of electrical insulation over the last 30 years may be summarized as follows: 1. The refinement of the theories which has resulted in a better understanding of dielectric phenomena. 2. The development of an imposing array of new synthetic materials which have replaced many of the natural materials formerly used. 3. The development of a new philosophy of thermally rating insulating materials and systems of materials as they are used in electrical equipment. 4. The application of the new materials in electric equipment in a manner which often results in major advances in equipment designs both from an economic and performance standpoint.

Because of space limitations the review presented in this paper is necessarily brief and only a few examples can be cited. It should be apparent from these, however, that the record of achievement over the past 30 years is an imposing one. It was undoubtedly accelerated through the co-operative efforts of physicists, engineers and chemists. The future looks even more promising. With the coming of the space age, insulation for electric equipment to operate at 500 C and higher in high-speed aircraft and rockets is already under development. The all-inorganic systems now being developed show considerable promise of finding applications other than in the military categories.

A considerable number of references are cited in this review, but these are far from complete. For a more complete reference to the literature in this field, the reader is referred to the "Digest of Literature on Dielectrics."[47] The National Research Council Conference on Electrical Insulation has performed a most valuable service in issuing this digest on an annual basis for more than 20 years.

References

1. The Modern Theory of Solids, F. Seitz, R. P. Johnson. *Journal of Applied Physics*, New York, N. Y., vol. 8, 1937, pp. 84, 186, 246.

2. Basic Processes of Gaseous Electronics (book), L. B. Loeb. University of California Press, Berkeley, Calif., 1955.

3. Dielectric Materials and Applications (book), A. R. von Hippel, editor. Massachusetts Institute of Technology Press, Cambridge, Mass.; also John Wiley and Sons, Inc., New York, N. Y., 1954.

4. Electric Breakdown of Perfluorocarbon Vapors and Their Mixtures With Nitrogen, L. J. Berberich, C. N. Works, E. W. Lindsay. *AIEE Transactions*, vol. 74, pt. I, *Communication and Electronics*, Nov. 1955, pp. 660–66.

5. The Dielectric Behavior of Some Fluorogases and Their Mixtures, G. Camilli, T. W. Liao, R. E. Plump. *Ibid.*, pp. 637–42.

6. Electric Breakdown of Solid and Liquid Insulation, A. R. von Hippel. *Journal of Applied Physics*, New York, N. Y., vol. 8, 1937, p. 815.

7. The Electrical Breakdown of Rocksalt at High Temperatures, L. Inge, A. Walther. *Technical Physics of the USSR*, Moscow, USSR, vol. 5, 1938, pp. 335–54.

8. Electrical Properties of Ceramics as Influenced by Temperature, E. W. Lindsay, L. J. Berberich. *AIEE Transactions*, vol. 67, pt. I, 1948, pp. 734–42.

9. Electric Strength and Molecular Structure of Saturated Hydrocarbon Liquids, R. W. Crowe, A. H. Sharbaugh, J. K. Bragg. *Journal of Applied Physics*, New York, N. Y., vol. 25, 1954, p. 1480.

10. The Triple Alliance in Electrical-Insulation, L. J. Berberich. *Electrical Engineering*, vol. 59, 1940, pp. 23–32.

11. Dielectric Behavior and Structure (book), C. P. Smyth. McGraw-Hill Book Company, Inc., New York, N. Y., 1955.

12. The Dielectric Strength of Insulating Fluids, Pt. I, Electric Strengths of Certain Gases and Gas-Vapor Mixtures, E. E. Charleton, F. S. Cooper. *General Electric Review*, Schenectady, N. Y., vol. 40, 1937, pp. 438–42.

13. The Use of SF_6 for High-Power Arc Quenching, W. M. Leeds, T. E. Browne, Jr., A. P. Strom. *AIEE Transactions*, vol. 76, pt. III, *Power Apparatus and Systems*, Dec. 1957, pp. 906–09.

14. Oxidation Inhibitors in Electrical Insulating Oils, L. J. Berberich. *ASTM Bulletin*, American Society for Testing Materials, Philadelphia, Pa., vol. 149, 1947, p. 65.

15. Non-flammable Dielectric Organic Compounds, F. M. Clark. *Industrial and Engineering Chemistry*, Washington, D. C., vol. 29, 1937, p. 698.

16. Insulated Electrical Conductors, E. H. Jackson, R. W. Hall. *U. S. Patent 2,307,588*, 1943.

17. Thermal Life of Enameled Magnet Wire, AIEE Committee Report. *AIEE Transactions*, vol. 78, pt. I, *Communication and Electronics*, May 1959, pp. 91–102.

18. Relation Between Chemical Composition and Dielectric Properties of Ceramic Materials With Low Dielectric Losses, M. D. Rigterink, R. O. Grisdale, S. O. Morgan. *Journal*, American Ceramic Society, Columbus, Ohio, vol. 25, 1942, pp. 439–43.

19. The Properties of Glass (book), G. W. Morey. Reinhold Publishing Corporation, New York, N. Y., 1938.

20. The Dielectric Properties of the Rutile Form of TiO_2, L. J. Berberich, M. E. Bell. *Journal of Applied Physics*, vol. 11, 1940, pp. 681–92.

21. High Titania Dielectrics, E. Wainer. *Transactions*, Electrochemical Society, Baltimore, Md., vol. 89, 1946, pp. 331–56.

22. General Principles upon Which Temperature Limits Are Based in the Rating of Electrical Equipment. *AIEE No. 1*, Apr. 1925, revised 1957.

23. International Activities in the Temperature Classification of Electrical Insulating Materials, L. J. Berberich, K. N. Mathes. *AIEE Transactions*, vol. 77, pt. III, *Power Apparatus and Systems*, Apr. 1958, pp. 69–73.

24. Guiding Principles in the Thermal Evaluation of Electrical Insulation, L. J. Berberich, T. W. Dakin. *Ibid.*, vol. 75, pt. III, *Power Apparatus and Systems*, Aug. 1956, pp. 752–61.

25. Loading Transformers by Temperature, V. M. Montsinger. *Ibid.*, vol. 49, Apr. 1930, pp. 776–92.

26. Stabilization of Chlorinated Diphenyl in Paper Capacitors, L. J. Berberich, R. Friedman. *Industrial and Engineering Chemistry*, vol 40, 1948, pp. 117–23.

27. Electrical Insulation Deterioration Treated as a Chemical Rate Phenomenon, Thomas W. Dakin. *AIEE Transactions*, vol. 67, pt. I, 1948, pp. 113–22.

28. Thermal Aging Properties of Cellulose Materials, G. Mämlow. *Acta Polytechnica*, Royal Swedish Academy of Engineering Sciences, Stockholm, Sweden, vol. 2, no. 1, 1949, pp. 1–67.

29. Thermal Evaluation of Enameled Magnet Wire, J. F. Dexter. *AIEE Transactions*, vol. 75, pt. III, *Power Apparatus and Systems*, Apr. 1956, pp. 40–44.

30. Test Procedure for Evaluation of the Thermal Stability of Enameled Wire in Air. *AIEE Standard no. 57*, Jan. 1959.

31. New Performance Standards for Elec-

trical Insulation of Rotating Machines, **AIEE Committee Report.** *AIEE Transactions*, vol. 73, pt. III-B, *Power Apparatus and Systems*, Dec. 1954, pp. 1542–46.

32. Test Procedure for Evaluation of Systems of Insulating Materials for Random-Wound Electric Machinery. *AIEE Standard no. 510*, Nov. 1956.

33. Test Procedure for Evaluation of Systems of Insulating Materials for Electric Machinery Employing Form-Wound Pre-Insulated Coils. *AIEE Publication no. 511*, Oct. 1956.

34. Proposed Test Code for Functional Temperature Endurance Tests of Ventilated Dry-Type Power and Distribution Transformers, **AIEE Committee Report.** *AIEE Transactions*, vol. 74, pt. III, *Power Apparatus and Systems*, Oct. 1955, pp. 971–77.

35. Test Procedure for Thermal Evaluation of Ventilated Dry-Type Power and Distribution Transformers. *AIEE Publication no. 65*, Nov. 1956.

36. Recommendations for the Classification of Materials for Insulation of Electrical Machinery and Apparatus in Relation to Their Stability in Service. *Publication no. 85*, International Electrotechnical Commission, Geneva, Switzerland, 1957.

37. A New High-Voltage Insulation for Turbine-Generator Stator Windings, **C. M. Laffoon, C. F. Hill, Graham Lee Moses, L. J. Berberich.** *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 721–30.

38. Silco-Flex Insulation, **J. L. Kuehlthau, P. A. Kryder.** *Allis-Chalmers Electrical Review*, Milwaukee, Wis., vol. 20, 1955, pp. 4–7.

39. An Advanced Concept for Turbine-Generator Stator-Winding Insulation, **E. J. Flynn, C. E. Kilbourne, C. D. Richardson.** *AIEE Transactions*, vol. 77, pt. III, *Power Apparatus and Systems*, June 1958, pp. 358–71.

40. Electrical Insulation—Its Application to Shipboard Equipment (book), **G. L. Moses.** McGraw-Hill Book Company, Inc., New York, N. Y., 1951.

41. Insulation Engineering Fundamentals (book), **G. L. Moses.** Lake Publishing Company, 1958.

42. The Application of a New Nontracking Butyl to High-Voltage Instrument Transformers, **R. A. Pfuntner, R. S. Norman, B. W. Walterdink.** *AIEE Transactions*, vol. 74, pt. III, *Power Apparatus and Systems*, Dec. 1955, pp. 1173–76.

43. Development and Application of Flexible Polyester Molding Materials for Instrument Transformers, **H. F. Minter, R. R. Kelley, A. J. Palumbo.** First National AIEE-NEMA (National Electrical Manufacturers Association) Conference on the Application of Electrical Insulation, Cleveland, Ohio, Sept. 1958 (available from the authors).

44. Transformer Cooling by Evaporation. *Westinghouse Engineer*, Pittsburgh, Pa., vol. 11, 1951, pp. 166–68.

45. Gas Insulated Transformers. *General Electric Review*, Schenectady, N. Y., vol. 59, 1956, pp. 41–44.

46. Power Capacitors (book), **R. E. Marbury.** McGraw-Hill Book Company, Inc., 1949.

47. The Digest of Literature on Dielectrics. National Research Council Conference on Insulation, National Academy of Sciences, Washington, D. C. (annual).

Remotely Controlled Electroluminescent Totaling Display

R. C. LYMAN
ASSOCIATE MEMBER AIEE

C. I. JONES
ASSOCIATE MEMBER AIEE

A. LEGER
NONMEMBER AIEE

THE NEED for automatic visual display of numerical information has grown rapidly in recent years. As a result, at least a dozen different numerical indicators are now available. Each will show any one of the ten Arabic numerals when actuated by the proper pattern of electrical signals. Some show painted numerals by electromechanical means; others use optical projection, masked lamps, edge-lighted plastic sheets, matrices of incandescent or neon lamps, numerically shaped wires in neon lamps, or cathode-ray tubes to indicate the desired numerals. Some distribute the ten numerals along a vertical line; some stack the numerals in depth, and some show each and every numeral in the same space and on the same plane. The device described here uses an electroluminescent phosphor to form a truly planar numerical indicator.

The applications for these numerical indicators are many, but they can be grouped conveniently into two classes. The first class contains numerical indicators used as substitutes for "scale and pointer" indicators in the meter and instrument field. Digital voltmeters, for example, are being used more and more frequently in laboratory and production measurements. The advantages of numerical display include better readability with no interpolation, parallax, or scale-factor errors, and more efficient use of panel-board area (a 2-digit indicator needs far less front panel area than a good 1% meter). The second class of applications concerns the replacement of manual posting of tabular data by automatic numerical display boards. This class is illustrated by the plans of several stock-exchange associations for automation of their entire operation, including the BID and ASK display boards, and by the military's requests for automatic numerical display boards to replace their chalk or grease pencil-on-acetate STATUS and ACTION boards. One advantage of automatic displays is faster and more accurate posting of changes in situation.

Most of the present automatic numerical display boards, which are typical of the second class of applications, use electromechanical indicators of the rotating belt or cylinder type. This paper describes a digital-display system of that class. However, it uses a new kind of solid-state indicator, an electroluminescent (EL) device,[1] for the display of numerical and symbolic information. The accompanying circuitry is composed of solid-state elements throughout. Thus the entire system is static; it is compact, requires low power, and offers high reliability.

Description of the System

Fig. 1 shows the digital-display system which is the subject of this paper. Such a system could be used by the dispatcher of a fleet of trucks, for example, to keep track of the availability and activity of his trucks, stationed at a distant location. The toggle-switch unit on the left of Figs. 1(A) and (B) would be located remotely, requiring only the single pair of wires running to the dispatcher's office. Each position on the display would correspond to a particular truck and would show if it was ready for assignment. The display would also show the total number of trucks available at that time.

The system has two functions: it reproduces the input information on the display board, and it summarizes that information. The encoding unit on the left in Fig. 1 has nine toggle-switch inputs and accompanying pilot lamps. The decoding unit receives the input information over the single pair of wires and controls the adjacent EL-display unit. The display unit shows the toggle-switch positions by a column of nine ON-OFF symbols (letter *A*s), and also shows the total number of ON switches

---

Paper 59-69, recommended by the AIEE Solid State Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 10, 1958; made available for printing November 28, 1958.

R. C. Lyman, C. I. Jones, and A. Leger are with Westinghouse Electric Corporation, Pittsburgh, Pa.

The authors wish to acknowledge the contributions by Dr. R. L. Bright to the synthesis of the system and the design of many of the transistor circuits.

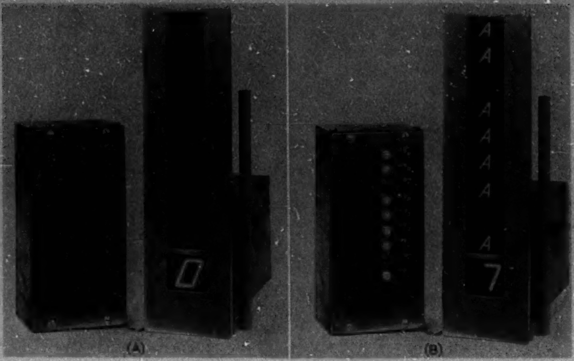


Fig. 1. Digital display system

by a numerical indicator at the base of the column. Because this device was intended primarily to illustrate the use of EL indicators, and not for installation in a particular system, only one short column was constructed. By direct expansion of the circuitry and display panel, a multicolumn board could readily be built; by simple extension of the method, the columns could each consist of many numerical indicators rather than ON–OFF (binary) indicators, with a multidigit summary at the base of each column.

In the paragraphs that follow, the construction of the display board, the operation of the system, and some of the circuits used will be described.

Construction of the Display Board

The binary and numerical indicators used in this system are self-luminous, using the phenomena of electroluminescence as their source of light. Just as visible light can be produced by exciting certain phosphors with ultraviolet radiation (photoluminescence) or with an electron beam (cathodoluminescence), so also visible light can be produced by exciting certain phosphors with an alternating electric field (electroluminescence).[2] Fig. 2 shows the construction of such an EL plate. The phosphor is mixed with a plastic dielectric and is sandwiched between a glass plate on one side and a metallic coating on the other side. The glass plate has a conductive but transparent coating on the side adjacent to the dielectric. The metallic coating is currently an evaporated layer of aluminum. If an alternating voltage is applied between the conductive transparent coating and the metallic layer of this "condenser," light is generated and will radiate through the glass plate.

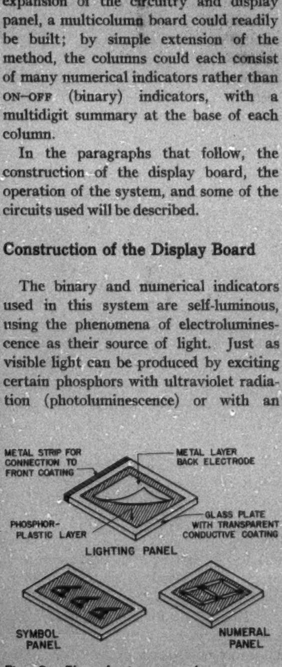

METAL STRIP FOR CONNECTION TO FRONT COATING
METAL LAYER BACK ELECTRODE
PHOSPHOR-PLASTIC LAYER
GLASS PLATE WITH TRANSPARENT CONDUCTIVE COATING
LIGHTING PANEL
SYMBOL PANEL
NUMERAL PANEL

Fig. 2. Electroluminescent plate construction, back view

For the display of symbols and numerals, the back electrode is evaporated through a stencil, with the results shown in the lower left and right parts of Fig. 2. The letter *A* symbols are excited by applying an alternating voltage between the conductive transparent coating and the segments forming the desired *A*. As is now apparent, the symbol *A* used in this system is an arbitrary choice; any symbol or set of symbols adaptable to stenciling could be used. As an alternative to the ON–OFF *A* symbol, the system could have used a double symbol such as *HK*, showing *H* for the switch ON and *K* for the switch OFF. A thin colored sheet of plastic, having good transmission in the blue and green wavelengths only, placed in front of the EL plate is especially desirable. It offers marked improvement in the contrast between the excited segments and the surrounding background.

The pattern used to show each of the ten Arabic numerals in the same area requires nine separate segments, as shown at the lower right of Fig. 2. Fig. 3 shows how excitation of the proper EL segments can show all ten numerals in easily recognizable form. The segment pattern shown evolved from a basic 7-segment figure-8 pattern and is only one of several possible elaborations.

The advantages of this type of self-luminous numerical and symbolic indicator are many. Some are related to the solid-state nature of electroluminescence. It requires no evacuated envelope or heated filament. As a result, the possibility of a catastrophic failure, as is common with incandescent lamps, is exceedingly small. This feature is of special importance where numerals are formed by many separately lighted segments; failure of a single segment can result in misreading some numerals. Because the EL plate is thin (essentially the thickness of the glass), it forms a truly planar indicator and therefore is readable over an exceedingly wide viewing angle. As a result of the method of construction, the plates can be made in a wide range of sizes and the numerals could range from several feet to a fraction of an inch in height. This use of electroluminescence has an additional advantage over the use of other light sources. Because it is an area source of light, and because only the desired shape need be excited, all the light generated is useful, none is wasted. Although projection systems may have comparable total light efficiency (expressed in lumens per watt), they require reflectors, lenses, and masks to form numerals and symbols. As a result, only a fraction of their total light output is available at the front of the indicator. For the same brightness of indication projection devices will dissipate much more heat than the EL area source, and heat is an important consideration in a large display board. Brightness of present EL indicators is sufficient for operation in adequately illuminated rooms. Future advantages include the

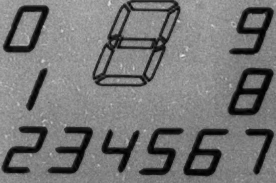


Fig. 3. Nine-segment numeral pattern

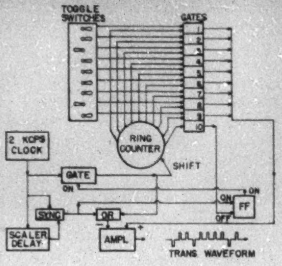


Fig. 4. Control and encoding circuit

property of change in emitted color with change in excitation frequency.

Description of System Operation

In this system, information is transmitted from the encoding unit to the decoding unit and display board over a single pair of wires in a series of pulse bursts. Each pulse burst contains the entire information to be displayed, making the system highly redundant. However, this results in rapid error correction should the information be garbled in transmission. The pulse bursts are obtained by sweeping the input circuit and interrogating each switch. If the switch is ON, a positive pulse results; if the switch is OFF, a negative pulse is generated. Although the system described here is entirely binary, decimal information may be encoded in a binary code, thus the binary design is not a system limitation. When the pulse burst is received at the decoding unit, the first pulse (which is always negative) resets the entire display to its OFF or zero state, as shown in Fig. 1(A). A counter then sums the positive pulses in the burst and causes the total to be displayed. The pulses are also distributed to memory circuits where the positive pulses cause the proper *A* symbols to be excited.

Fig. 4 shows the block diagram of the circuitry used to obtain the pulse bursts. To describe the operation of the circuitry, the flip-flop is assumed to be initially ON. The clock runs at a 2,000-cps (cycles per second) rate. Because the flip-flop is ON, the clock gate conducts pulses to the ring counter, causing it to successively actuate each of the ten gate circuits. If a particular toggle switch is thrown OFF, no pulse will be generated at the corresponding gate when the ring counter actuates that gate. Hence, the only pulse appearing at the output amplifier will be the clock pulse conducted through the clock gate and the OR circuit. It will cause a negative-output pulse to be transmitted. If a toggle switch is thrown ON, the coincidence of the ring counter and toggle-switch signals will generate a pulse at the respective gate. This pulse goes to the positive input of the amplifier and, by inhibiting the clock pulse appearing at the negative input, causes a positive-output pulse to be transmitted. This operation continues until the ring counter has scanned all nine switches and reaches the tenth gate. It then turns the flip-flop OFF, thus disabling the clock gate, and ending the transmitted pulse burst.

Clock pulses are also used to drive a scaler delay, which sets the repetition rate of pulse bursts. The scaler has a capacity of about 50 counts; it generates an output pulse, resets, and begins scaling again after each 50th clock pulse. Because its output pulse comes just after the clock pulse, it is synchronized with the next clock pulse. It starts the transmission of another pulse burst by causing a negative pulse to be transmitted and by turning the flip-flop ON again. The flip-flop then opens the clock gate and the ring counter again scans the input switches, beginning on the next clock pulse. Circuit element delays are sufficient to prevent this operation from occurring at the same clock pulse. The negative pulse is followed, therefore, by nine information pulses. The burst repetition rate is 1/50th of the clock frequency, i.e., 40 cps.

The decoding and display circuitry is shown in Fig. 5. The first incoming pulse of the burst, being negative, triggers the monostable multivibrator. This in turn resets the ring counter, the summation counter, and the memory flip-flops, and also blanks the numerical display for the duration of the pulse burst. The remaining nine incoming pulses serve three purposes: 1. They are rectified and used to drive this ring counter synchronously with the encoding ring counter. 2. The positive pulses appear at the gate circuits and the coincidence of any positive pulse with its corresponding ring-counter output produces an output pulse at the respective gate. This output pulse turns the flip-flop ON and the *A* symbol corresponding to the proper toggle switch is excited by the control circuit. 3. The positive pulses also appear at the summation counter where they are totaled. The state of the counter is translated to the required 9-segment code and actuates control of the numerical display. This translation is accomplished by an AND–OR diode matrix. Because the output of the translation matrix indicates zero when the summation counter is reset and then indicates each successive number as the counter totals the positive pulses, the display is blanked until the transmission is completed. The blanking is not readily noticeable because the burst repetition rate is 40 cps, and because the transmission lasts only one-fifth of the 25-millisecond repetition period.

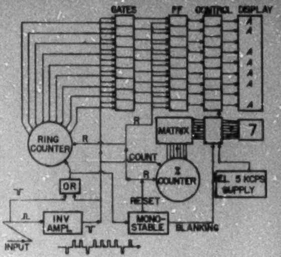


Fig. 5. Decoding and totaling circuit

Although a pair of wires are used for transmission in this system, a single channel of any kind, having sufficient bandwidth to pass the pulse bursts, would serve equally well. Slower transmission rates would permit transmission over narrower bandwidth channels. Applications suggested by these comments include binary, numerical, or symbolic indicators for supervisory control and other telemetry systems. The manually operated toggle switches of this system could be replaced by any type of automatic switching device. The decoding and display unit could serve as a special purpose output device for a digital computer, provided the proper pulse pattern could be generated.

Description of the Circuitry

Because the circuit used for control of the EL segments was designed especially for this application, and has not been described elsewhere in the literature, it is described in some detail in the following. The scaler-delay circuit combines magnetic cores with transistors and is explained briefly. The counters and the translation matrix are also described. The rest of the circuits used in this system are essentially standard transistor-circuit designs and are not included in this paper.

All the circuits for this system are constructed on etched cards, for economy of space and convenience of assembly. The advantage of this method is illustrated by the following example. Nine flip-

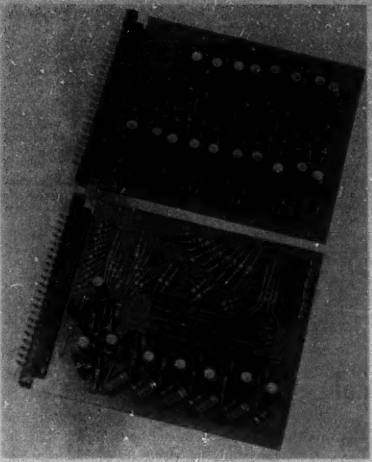


Fig. 6 (left). Flip-flop and ring counter cards

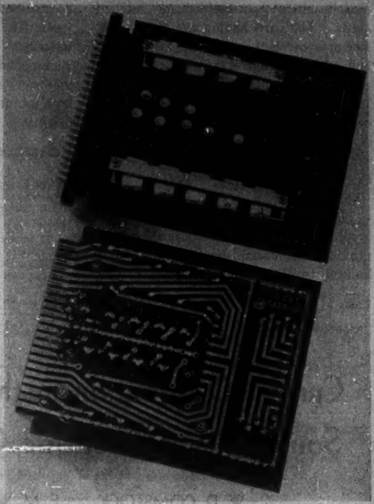


Fig. 8 (right). Electroluminescent-display control circuit card

flops are required to store the states of the *A* symbols. They require 18 transistors, which are all assembled on one 4-by-5-inch 22-contact card. Fig. 6 shows the component sides of that card, and a ring counter-with-gates card. A total of three cards are used for the encoding circuitry, and six cards for the decoding and totaling circuitry.

Counter Circuits

The same circuit is used for both ring counters and for the summation counter. A 10- or 9-state device is required in each case. The circuit departs from common practice in that a flip-flop driving a ring of five is used rather than a ring of ten or a 4-stage binary counter. A ring of ten requires ten transistors and places rather stringent requirements on them because each base must couple to the other nine collectors. The flip-flop and ring-of-five combination uses only eight transistors, including a buffer between the flip-flop and the ring circuit. As a result, each of the nine gates in the encoding and the decoding units is a 3-input diode AND, having a switch or transmission input, an input from one side of the flip-flop, and an input from one stage of the ring counter. Coincidence of pairs of the last two inputs determines each of the ten states.

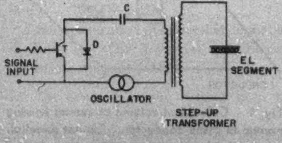


Fig. 7. Electroluminescent-display control circuit

Translation Matrix

Another result of using the flip-flop and ring-of-five combination is that translation from the state of the summation counter to the 9-segment code is neither straight binary nor decimal. The method of Karnaugh Maps[3,4] was adapted to fit this unusual input code and was used to minimize the number of diodes required in the AND–OR matrix.

Scaler Delay

Rectangular loop magnetic cores and switching transistors are combined in the scaler-delay circuit. The scale factor of about 50 is accomplished by one pulse-generator core and one counting core, together with three transistors. The circuitry is similar to that described at the 1955 National Electronics Conference[5] and in other references.[6,7] It makes use of the high-remanence flux of such magnetic materials.

Display Control Circuit

The electroluminescence described in this paper is an alternating-field phenomenon. Typical operation for display purposes is at a frequency of 400 cps and an alternating potential of 230 volts rms. Some means of control is needed to gate this high-frequency high-voltage power to the EL segments in response to the low direct-voltage logic signals. A single transistor a-c switch was developed for this purpose. Such a switch has three basic requirements; it must pass current efficiently in either direction when closed, it must block voltage of either polarity when open, and it must have a good ON to OFF resistance ratio. Most junction transistors can pass current efficiently in only one direction and can block voltage effectively from only one direction. By adding a shunt diode from collector to emitter, and a capacitor in series with this transistor-diode combination, all requirements of an a-c switch are met; see Fig. 7. When the transistor is biased ON, currents from emitter to collector pass through the transistor while currents from collector to emitter pass through the diode. The capacitor, if properly chosen with respect to the load, will offer negligible impedance. When biased OFF, the transistor and diode both block during the half cycle when the instantaneous polarity is positive at the emitter. On opposite half cycles, the capacitor charges through the diode to the peak value of the supply voltage and no further current flows. Although the transistor and diode must each block twice the peak supply voltage, the load current is divided between the two. The step-up transformer is required because presently available transistors do not have the necessary voltage rating. Fig. 8 shows these a-c switch circuits assembled on etched cards. The card having the component side up controls the numerical display and has only eight control circuits. Although the display is composed of nine segments, Fig. 3 shows that one segment is always ON and hence does not require active control.

Conclusions

The most important feature of this system is the use of electroluminescence to provide the visual display. It is not limited to the circuitry here described nor to the few applications mentioned before. The marked advantages of EL numerical and symbolic indicators include: solid-state construction, wide-angle viewing, wide range of size and shape, high useful luminous efficiency, and good brightness.

References

1. AN ELECTROLUMINESCENT DIGITAL INDICATOR WITH ELPAK TRANSLATION LOGIC, **E. A. Sack.** *AIEE Transactions*, vol. 77, pt. I, Mar. 1958, pp. 113–18.

2. ELECTROLUMINESCENCE AND RELATED TOPICS, **G. Destriau, H. F. Ivey.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 43, Dec. 1955, p. 1911–40.

3. THE MAP METHOD FOR SYNTHESIS OF COMBINATIONAL LOGIC CIRCUITS, **M. Karnaugh.** *AIEE Transactions*, vol. 72, pt. I, Nov. 1953, pp. 593–99.

4. ANALYZING COMBINATIONAL CIRCUITS BY BOOLEAN MATRICES AND KARNAUGH MAPS, **B. Beizer, S. W. Leibholz.** *Electrical Manufacturing*, New York, N. Y., June 1958, p. 98 ff.

5. A PREDETERMINED SCALER UTILIZING TRANSISTORS AND MAGNETIC CORES, **R. I. Van Nice, R. C. Lyman.** *Proceedings*, National Electronics Conference, Chicago, Ill., vol. XI, 1955, pp. 859–69.

6. PREDETERMINED SCALER UTILIZES TRANSISTORS AND MAGNETIC CORES, **R. I. Van Nice, R. C. Lyman.** *Electronic Equipment*, New York, N. Y., Feb. 1956, p. 46 ff.

7. A HIGH-ACCURACY STATIC TIME DELAY DEVICE UTILIZING TRANSISTORS, **G. F. Pittman, Jr.** *AIEE Transactions*, vol. 74, pt. I, Mar. 1955, pp. 54–58.

Current Distribution in the Cylindrical Source Plane-Electrode Configuration

W. D. COMSTOCK
ASSOCIATE MEMBER AIEE

E. M. WILLIAMS
MEMBER AIEE

THERE ARE a number of instances in the operation of electric apparatus in which the physical situation may be approximated by a configuration in which a plane electrode conducts current which is received from or delivered to a cylindrical conducting path normal to its surface. Among these instances the most familiar examples are cases in which the conducting path is an ionized column resulting from sparkover (breakdown) of a solid, liquid, or gaseous dielectric between conductors, or the process of arc formation by the separation of initially juxtaposed conducting surfaces between which current is flowing. Typical situations correspond to circumstances accompanying insulation failures in apparatus, relay contact "breaks," circuit-breaker operations, and to the sparkover phenomena utilized in electric spark machining.[1,2] In each of these situations craters are formed on the electrode surface. Considerable attention[3–5] has been given to the phenomena of crater formation for obvious reasons. Formation of the craters has been attributed to various causes[1,5,6] and, indeed, the predominant cause may vary, depending on conditions of current formation, duration and magnitude, and on current-density distribution in the ionized column. In at least one range of physically realizable circumstances current density beneath the surface has been demonstrated[3] to be significant in determining the depth of crater formed by a particular discharge. The relationship demonstrated was of a qualitative character only, however, owing to the lack of quantitative knowledge concerning both transient and steady-state subsurface current distribution. This paper describes the results of an analytical study of current distribution under appropriate conditions. In this study steady-state current distribution was determined from work by others on an analogous problem in hydrodynamics.[7,8] Transient current distribution was determined by a solution which is described in some detail in the Appendix.

The problem of interest was reduced to the following:

> A cylindrical current column of uniform current density instantly starts flowing, normal to the surface of a plane conductor. The diameter of the column is assumed constant. What is the current distribution beneath the surface of the conductor and how does it vary with time?

The solution of this problem in the form stated should be of direct value in the study of electrode phenomena following sparkover in gas dielectrics, in which the discharge cross section is believed to remain relatively constant as a function of time. However, in liquid dielectrics the discharge cross section increases in area approximately linearly with time over a wide range of time duration[3] during a transient interval before quasi-stable pressure and temperature conditions are reached. Nevertheless, in the time interval in which area is increasing the solution of the above problem can be used advantageously to estimate the conditions under which skin-effect will be important. Moreover this solution should apply well to the relatively long duration (in terms of the time intervals in which skin-effect is important) discharges in the circuit breaker.

It should be recognized that the progressive formation of enlarging craters at the termini of an actual discharge would modify the physical situation in a way that is not taken into account by the analytical model suggested. Owing to

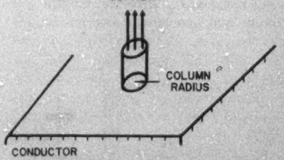


Fig. 1. Cylindrical column of current flowing normal to plane conductor. Current direction may be either into or out from conductor; mathematical analysis based on the latter

---

Paper 59-46, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted May 13, 1958; made available for printing November 21, 1958.

W. D. COMSTOCK is with Syracuse University Research Corporation, Syracuse, N. Y., and E. M. WILLIAMS is with Carnegie Institute of Technology, Pittsburgh, Pa.

The work reported here was supported by the National Science Foundation under Grant *G3072*. This paper comprises part of a dissertation submitted by W. D. Comstock in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Carnegie Institute of Technology.

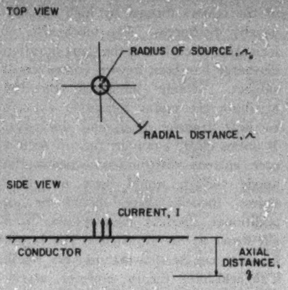


Fig. 2. Mathematical model of source of uniform current density on a conducting infinite half-plane

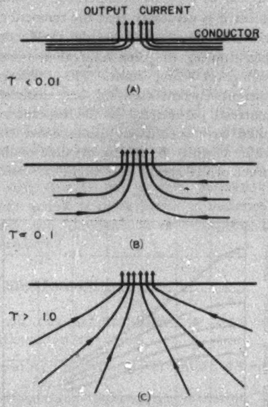


Fig. 3. Current distribution in a plane electrode under and adjacent to cylindrical current source

A—Extreme skin-effect range
B—Intermediate skin-effect range
C—Steady-state condition

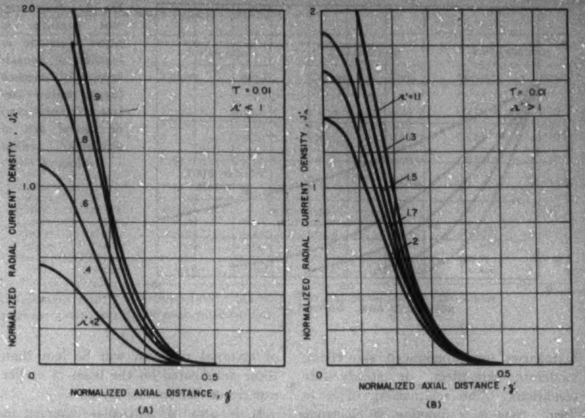


Fig. 4. Typical radial current densities in electrode plane (under discharge terminus) as function of normalized axial distance 3 feet below surface of electrode, at various normalized radii, r', from center of discharge, in extreme skin-effect range

the present absence of knowledge concerning the dynamics of formation of such craters this difficulty is unavoidable. However in the discussion of the results of this investigation it will appear that this solution nevertheless permits some deduction as to the rate of crater formation.

The configuration of interest is depicted in Fig. 1. The following additional conditions are postulated:

1. The diameter of the current path is constant.

2. The current density in the current column is uniform throughout its cross-section. This is actually not the case except, possibly, for an initial condition following sparkover, but no information is available as to actual transient current-density distribution.

3. The current is established instantaneously at an initial value which is sustained throughout the period of interest. Actually a finite time interval is required for establishment of current in sparkover cases; during this interval current rises and voltage falls. For very short gaps this interval is known[3] to be of the order of 20 millimicroseconds (mμs). Certainly, as far as microsecond and longer phenomena are concerned the duration of the initial interval may be neglected.

A mathematical model of the physical situation is depicted in Fig. 2. It comprises a source of uniform current density acting over a circular area on the surface of an infinite half-plane of conducting material.

The process of analytical solution is summarized in the Appendix. The resulting computed characteristics of the current distribution are described by normalized radial and axial components of the current density, $J_r' = J_r/J$ and $J_z' = J_z/J$ respectively, as functions of the normalized space parameters, $z' = z/r_0$ and $r' = r/r_0$, and the normalized time parameter, $t/\mu\sigma r_0^2$. J_r and J_z are actual current densities and J is the current density in the discharge column. The depth below surface is z, r is radial distance from center of current column, r_0 is radius of current column, t is time in seconds after initiation of the current, μ is meter-kilogram-second (mks) magnetic permeability of conductor, and σ is (mks) electrical conductivity of conductor.

The physical nature of the results is illustrated in Fig. 3. Three ranges of τ are defined as: $\tau < 0.01$, extreme skin-effect range; $0.01 < \tau < 1.0$, intermediate skin-effect range; and $\tau > 1.0$, steady-state range.

Results of considerable generality are obtained as a consequence of the use of normalized quantities. In the extreme skin-effect range, the radial component of current density is given by the simple asymptotic expressions of equations 14 (A) and 14(B). Typical values of J_r' for $\tau = 0.01$ are plotted in Fig. 4. Axial current densities, on the z axis, which amounts to an extension of the axis of the discharge, are given by expressions of equations 15(A) and 15(B) and are included among the curves of Fig. 5. In the intermediate skin-effect range, equation 11 for the radial current density has been evaluated for a number of values of τ by the use of a digital computer. Typical values are plotted for $\tau = 0.1$ in Fig. 6. Values of steady-state radial and axial current densities are shown in Figs. 7 and 5, respectively.

Although the results given for axial current density are restricted to $r' = 0$ much useful information can be obtained; in particular, these values are applicable in the estimation of the central depth of craters to be produced by discharges.[1-3]

It is of considerable interest to determine the relationship of τ to time in seconds and the corresponding time

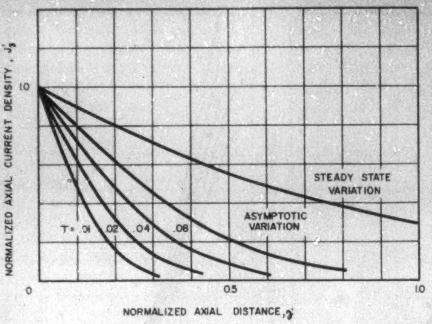


Fig. 5. Axial current densities in plane electrode as function of normalized axial distance (depth below surface of conductor) on central axis of discharge; various values of τ

boundaries for pronounced skin-effect, moderate skin-effect, or steady-state conditions. The relationship of τ to "real" time is a function of electrode permeability and conductivity, and of the discharge current and current density. Table I lists some values of real time corresponding to values of τ for two current densities and two electrode materials. The first value of current density, 10,000 amp (amperes) per square inch, is of the order of magnitude of current densities in sustained arcs in gases at atmospheric pressure; in such arcs current density is constant with time. The second, very high, value of current density is typical[2,3] of initial transient values following sparkover in a liquid dielectric. Here progressively growing cross-sectional areas and, correspondingly, progressively decreasing current densities are always involved so that the values of τ listed are useful only in estimating the extent of skin-effect. However, as a means of arriving at a quick decision as to skin-effect, or no skin-effect, the data are useful.

For example, there is considerable interest at present in the development of electric spark machine tools for the forming of large aircraft dies, involving working gaps of the order of several square feet or more in cross sections. Pulse currents of the order of 10,000 amps with durations of not more than 50 μsec (microseconds) have been suggested as design objectives although these currents are considerably in excess of those that have been used in the past and, indeed, in excess of currents for which there is any operating experience. It is clear from Table I that for such discharge currents and with electrodes of good conductors, or especially of ferromagnetic tool steels, there will be well-defined skin-effect in the electrode surfaces and, consequently, that the amounts of material removed will be less than those anticipated on the basis of earlier experience with smaller currents.

It is also of particular interest to examine the computed current densities for the ferromagnetic alloy and the nonferromagnetic alloy of otherwise similar mechanical properties in the experiment described by Williams and Smith[3] to show the influence of current density beneath an electrode surface on the depth of the craters formed by high current-density discharges. The mechanism of formation of craters in the short-duration discharge has been a very controversial subject. Widely advanced theories[6] attribute the crater formation entirely to heat transfer from the discharge. It seems clear, however, that in such a case current distribution beneath the anode surface would have no effect unless, indeed, there should be an additional thermal contribution from i^2R losses in the electrode. The latter, however, can be shown to be very small. The field-force theory, however, attributes crater formation primarily to axial current density, at least for short-duration high current-density discharges. In order for a difference in depth to appear as a result of difference in skin-effect it is necessary that the removal of material from an electrode shall take place during the time interval in which skin-effect is important. The discharge current, current duration, and electrode material parameters in the experiment cited were such that τ approaches the value of unity in an elapsed time of the order of 0.1 μsec. This suggests, there-

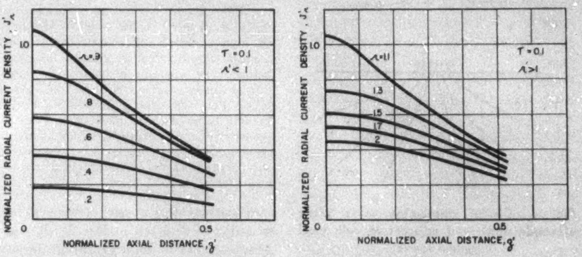


Fig. 6. Typical radial current densities in plane electrode (under discharge terminus) as function of normalized axial distance (depth below surface of electrode) at various normalized radii, r', from center of discharge in intermediate skin-effect range

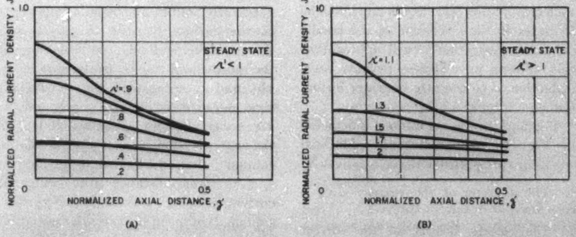


Fig. 7. Steady-state radial current densities in plane electrode with data as for Fig. 6

Table I. Values of τ for Some Typical Conditions*

Elapsed time		I=1 amp		I=10 amp		I=100 amp		I=1,000 amp		I=10,000 amp	
A†	B†	Copper	Steel	Copper	Steel	Copper	Steel	Copper	Steel	Copper	Steel
0.15 μsec	150 μμsec‡	0.1	1×10^{-3}	1×10^{-2}	1×10^{-4}	1×10^{-3}	1×10^{-5}	1×10^{-4}	1×10^{-6}	1×10^{-5}	1×10^{-7}
1.5 μsec	0.15 mμsec	1.0	1×10^{-2}	0.1	1×10^{-3}	1×10^{-2}	1×10^{-4}	1×10^{-3}	1×10^{-5}	1×10^{-4}	1×10^{-6}
15 μsec	1.5 mμsec	steady state	0.1	1.0	1×10^{-2}	0.1	1×10^{-3}	1×10^{-2}	1×10^{-4}	1×10^{-3}	1×10^{-5}
150 μsec	15 mμsec	steady state	1.0	steady state	0.1	1.0	1×10^{-2}	0.1	1×10^{-3}	1×10^{-2}	1×10^{-4}
1.5 ms§	0.15 μsec	steady state	steady state	steady state	1.0	steady state	0.1	1.0	1×10^{-2}	0.1	1×10^{-3}
15.0 ms	1.5 μsec	steady state	steady state	steady state	steady state	steady state	1.0	steady state	0.1	1.0	1×10^{-2}
150 ms	15 μsec	steady state	steady state	steady state	steady state	steady state	steady state	steady state	1.0	steady state	0.1
1.5 s‖	150 μsec	steady state	steady state	steady state	steady state	steady state	steady state	steady state	steady state	steady state	1.0

* For the copper electrode $\sigma=5.87\times10^7$ mho meters and $\mu=\mu\text{ohm}=4\pi\times10^{-7}$. For the steel electrode $\sigma=0.58\times10^7$ mho meters and $\mu=1{,}000$ μohms.

† Two discharge current densities are assumed: A—10^4 amps per square inch; B—10^8 amps per square inch.

‡ Micromicrosecond.

§ Millisecond.

‖ Seconds.

fore, that an appreciable amount of crater formation must have taken place in a time interval shorter than 0.1 μsec. This conclusion and the general character of experimental data reported by Williams[3] and Smith suggest that crater formation takes place continuously and progressively during a discharge. Although this is not a very surprising deduction it is nevertheless at variance with the physical process suggested elsewhere.[9]

Nomenclature

All physical quantities are given in rationalized mks units.

J = axial current density at surface of conductor
I = total current flowing out of conductor
$A=\pi r_0^2$, the area of current source
r_0 = radius of the area of current source
$\mathbf{E}$ = electric field intensity
$\mathbf{H}$ = magnetic field intensity
$\mathbf{B}$ = magnetic flux density
ρ = charge density per unit of volume
ϵ = electric permittivity of conductor
μ = magnetic permeability of conductor
σ = electric conductivity of conductor
t = time
r, ϕ, z = radial, angular, and axial co-ordinates respectively in cylindrical co-ordinate system
$\mathbf{J}$ = electric current density $=\sigma\mathbf{E}$
J_r, J_z = radial and axial components, respectively, of the current density
J_r', $J_z'=J_r/J$, J_z/J = normalized radial and axial current density respectively
$r'=r/r_0$ = normalized radial distance
$z'=z/r_0$ = normalized axial distance
$\tau=t/\mu\sigma r_0^2$ = normalized time parameter
J_{rss}', J_{zss}' = normalized steady-state radial and axial current density respectively

Appendix

Referring to Fig. 2, the axial current density of the source is

$$J=\frac{I}{A} \tag{1}$$

in which

I = total current flowing out of the conductor
$A=\pi r_0^2$, the cross-sectional area of the discharge

Maxwell's equations govern the electric and magnetic field variation in the conductor.

These are

$$\nabla\times\mathbf{E}=-\frac{\gamma\mathbf{B}}{\gamma t} \tag{2}$$

$$\nabla\times\mathbf{H}=\sigma\mathbf{E}+\epsilon\frac{\gamma\mathbf{E}}{\gamma t} \tag{3}$$

$$\nabla\cdot\mathbf{E}=\frac{\rho}{\epsilon} \tag{4}$$

$$\nabla\cdot\mathbf{B}=0 \tag{5}$$

Since the source is circular a cylindrical coordinate system is used. Let r, ϕ, and z represent the radial, angular, and axial coordinates respectively.

The boundary conditions for this physical situation are

$$\begin{aligned} J_z\Big|_{z=0} &= J,\ 0\leq r<r_0\\ &=0,\ r_0<r\leq\infty \end{aligned}$$

$$J_z=J_r=0,\qquad \sqrt{z^2+r^2}=\infty$$

The initial conditions are

$$J_z=J_r=0,\ t<0$$

$$J_z=J,\ \underset{\substack{z=0\\ t\geq0}}{0\leq r<r_0}$$

The net charge densities in the conductor are assumed to be negligible. For this case the charge density term, ρ/ϵ, in equation 4 as well as the displacement current density, $\epsilon\partial\mathbf{E}/\partial t$, in equation 3 may be neglected for this analysis because times are to be considered which are long relative to the relaxation time, ϵ/σ, of the conductor. ϵ/σ is of the order of 10^{-14} μsec for a good conductor. Under these assumptions equations 2 through 4 can easily be shown to reduce to two equations in terms of the current density.

$$\nabla^2J_z=\mu\sigma\frac{\partial J_z}{\partial t} \tag{6}$$

$$\nabla\cdot\mathbf{J}=0 \tag{7}$$

in which

∇^2 = Laplacian operator in cylindrical coordinates
$\nabla\cdot$ = divergence operator in cylindrical coordinates

Equations 6 and 7, together with the boundary conditions and initial conditions required for J_z, are sufficient for solutions for J_z and J_r. First, to solve for J_z, the Laplace transform of equation 6 is taken. The boundary conditions imposed on J_z are also transformed. The inverse Laplace transform of the resulting solution in the s domain yields a solution for J_z in the time domain. This solution for J_z is used to obtain the solution for J_r through equation 7.

The Laplace transform of equation 6 yields

$$\nabla^2J_z(s)=\mu\sigma sJ_z(s) \tag{8}$$

in which

$J_z(s)=\mathcal{L}[J_z(t)]$

The combination of boundary and initial conditions for J_z are

$$J_z\Big|_{z=0}=f(r)=\frac{J}{s},\ \underset{z=0}{0\leq r<r_0}$$

$$J_z=0,\quad \sqrt{z^2+r^2}=\infty$$

Solution of equation 8, subject to the boundary conditions, can be shown to yield

$$J_z=\frac{J_{r_0}}{s}\int_0^\infty J_1(k_{r_0})J_0(kr)e^{-\sqrt{k^2+\mu\sigma s}\,z}\,dk \tag{9}$$

in which

$J_0(u)$ = the Bessel function of first kind, order zero, with argument u
$J_1(u)$ = the Bessel function of first kind, order one, with argument u

The inverse Laplace transform of the solution in the s domain, represented by equation 9, can also be shown to yield

$$J_z' = \int_0^\infty J_1(k)J_0(kr')\cdot\frac{1}{2}\{\epsilon^{-kz'}\times [1+\mathrm{Erf}(y)]+\epsilon^{kz'}[1-\mathrm{Erf}(x)]\}\,dk \qquad (10)$$

in which

J_z', r', z', and τ = meanings given in the Nomenclature.
$x = z'/2\sqrt{\tau}+k\sqrt{\tau}$
$y = -z'/2\sqrt{\tau}+k\sqrt{\tau}$
$\mathrm{Erf}(u)$ = the error function with argument u

A solution for J_r can be effected through the use of equation 7. Solution of this equation can be shown to yield

$$J_r' = \int_0^\infty J_1(k)J_1(kr')\cdot\frac{1}{2}[\epsilon^{-kz'}\{1+\mathrm{Erf}(y)\}+ \epsilon^{kz'}\{1-\mathrm{Erf}(x)\}]dk + \frac{\epsilon^{-z'^2/4\tau}}{\sqrt{\pi\tau}}\int_0^\infty J_1(k)J_1(kr')\,\frac{\epsilon^{-k^2\tau}}{k}\,dk \qquad (11)$$

Equations 10 and 11 reduce in the steady state to

$$J_{z_{ss}}' = \int_0^\infty J_1(k)J_0(kr')\epsilon^{-kz'}dk \qquad (12)$$

$$J_{r_{ss}}' = \int_0^\infty J_1(k)J_1(kr')\epsilon^{-kz'}dk \qquad (13)$$

These steady-state solutions check with those given for an analogous problem in hydrodynamics.[7,8]

An asymptotic form for radial current density may be derived for small values of τ and values of r' sufficiently removed from $r'=1$. This can be shown to be

$$J_r' \approx \frac{\epsilon^{-z'^2/4\tau}}{\sqrt{\pi\tau}}\cdot\frac{r'}{2},\ r'<1 \qquad (14A)$$

and

$$J_r' \approx \frac{\epsilon^{-z'^2/4\tau}}{\sqrt{\pi\tau}}\cdot\frac{1}{2r'},\ r'>1 \qquad (14B)$$

The asymptotic form for J_r' can be substituted in equation 7 to obtain an asymptotic form for J_z' as

$$J_z' \approx \mathrm{Cerf}\left(\frac{z'}{2\sqrt{\tau}}\right),\ r'<1 \qquad (15A)$$

and

$$J_z' \approx 0,\ r'>1 \qquad (15B)$$

in which

$\mathrm{Cerf}(u)$ = the complementary error function with argument u

Evaluation of the axial current density has not been carried out in its entirety because of the complexity of numerical work involved. However, a comparison of the asymptotic form and the steady-state solution should permit an estimate of axial current density at intermediate values of τ. The steady-state axial current density at $r'=0$ is

$$\underset{r'=0}{J_{z_{ss}}'} = \int_0^\infty J_1(k)\epsilon^{-kz'}dk = 1-\frac{z'}{\sqrt{1+z'^2}} \qquad (16)$$

This function appears with the asymptotic curves of Fig. 5.

References

1. Theory of Electric Spark Machining, **Everard M. Williams.** *AIEE Transactions*, vol. 71, pt. II, Jan. 1952, pp. 105–08.

2. Recent Developments in the Theory and Design of Electric Spark Machine Tools, **E. M. Williams, J. B. Woodford, Jr., Richard E. Smith.** *Ibid.*, vol. 73, pt. II, May 1954, pp. 83–88.

3. Phenomena Accompanying Transient Low-Voltage Discharges in Liquid Dielectrics, I—Anode Phenomena at Low Currents, **E. M. Williams, R. E. Smith.** *Ibid.*, vol. 74, pt. I, May 1955, pp. 164–69.

4. High-Current Arc Erosion of Electric Contact Materials, **W. R. Wilson.** *Ibid.*, pt. III, Aug., pp. 657–64.

5. Mechanism of Electro-Erosion of Metals, **S. L. Mandelshtam, S. M. Raiskii.** *Izvestiya Akademii Nauk SSSR*, Moscow, USSR, Ser. Fiz., vol. 13, no. 5, 1949, pp. 549–65.

6. Physical Foundation of Electro-Spark Working of Metals, **B. N. Solotych.** *Schriftenreihe*, Verlages Technik, vol. 175, VEB Verlag Technik, Berlin, Germany, 1955.

7. Hydrodynamics (book), **H. Lamb.** Dover Publications, Inc., New York, N. Y., 1945, pp. 137–39.

8. An Application of Hankel Transforms in Axially Symmetric Potential Flow, **A. G. Machie.** *Proceedings*, Edinburgh Mathematical Society, Edinburgh, Scotland, vol. 9, July 1956.

9. The Electrospark Method of Machining of Metals, **B. R. Lazarenko.** *Vestnik Mashinostroeniya*, Moscow, USSR, vol. 27, no. 1, 1947, pp. 25–36.

Engineering Aspects of TASI

K. BULLINGTON
NONMEMBER AIEE

J. M. FRASER
NONMEMBER AIEE

FOR MANY YEARS telephone people have been intrigued with the idea of making use of the idle time during a telephone call. Time Assignment Speech Interpolation (TASI) is a high-speed transmission and switching system based on the principle of using the free channel time to interpolate additional talkers. Although TASI requires considerable terminal equipment, it can approximately double the usefulness of long expensive channels such as the deep sea submarine cable systems. The first speech interpolation system is expected to be placed in service during 1960 on the first transatlantic telephone cable system between London and New York, which was opened for service in September 1956.

The basic principles of TASI are old. An extensive investigation of the feasibility of speech interpolation systems was conducted in 1946 and 1947 by A. C. Dickieson, P. G. Edwards, and others. An experimental model was demonstrated in 1950 by A. E. Melhose, and this was later extended and improved by R. L. Carbrey.

In a normal 2-way conversation each talker ordinarily uses the circuit for only about half of the time. In addition, the circuit may not be used momentarily while the operator is trying to get the right people together, or while an individual comes to the phone or takes the normal pauses between sentences and even syllables. All of these gaps add up to a substantial amount of free time. Measurements on working circuits shown in Fig. 1 indicate that the average activity (percentage of time that energy above a very low threshold is transmitted in one direction) is no more than 35 to 40% of the time that the circuit is busy at the switchboard. Since long-distance circuits use separate pairs of wires or separate carrier channels for the two directions of transmission, it follows that, on the average, each 1-way channel is free for 60 to 65% of the time.

An attempt to interpolate two or more independent conversations on a single channel would create considerable delay and mutual interference, since the probability of two talkers wanting the same channel at the same time is moderately high. However, when 2 or 3 dozen channels are operated as a group, these variations in individual activity tend to average out and the probabilities can be counted on to minimize the mutual interference to a point where there will be no noticeable effect on continuity of conversation. For example, even when all the channels are busy, at least one channel on the average is available for reassignment every L/c seconds where L is the average talkspurt length (which is somewhat less than one second)

Paper **59-70**, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 31, 1958; made available for printing November 28, 1958.

K. Bullington and J. M. Fraser are with the Bell Telephone Laboratories, Inc., Murray Hill N. J.

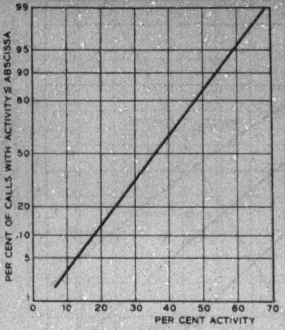


Fig. 1. Measured activity distribution on transatlantic calls

and c is the number of channels in the group.

A simplified block diagram for a TASI system is illustrated in Fig. 2 for one direction of transmission. An identical but independent arrangement is used for the opposite direction. Each of the n lines from the toll switchboard is equipped with a speech detector which is capable of recognizing the weakest speech within 5 milliseconds or less. When the speech detector on line 1 operates, the TASI switching network connects line 1 to an idle channel, say channel c, and then sends on that channel, a 10 to 15-millisecond coded signal which instructs the distant receiving terminal to connect that channel to listener line 1. Once a connection has been set up it is maintained until someone else needs a channel, even though the person may stop talking in the meantime. A similar 10- to 15-millisecond coded signal is used for disconnect purposes, but this is sent over a separate channel reserved for this purpose in order to avoid interference to the listener.

When speech appears on any of the n lines, the speech must be recognized by the speech detector, which causes some clipping during the first few milliseconds. This type of clipping can be minimized by increasing the sensitivity of the speech detector. The sensitivity cannot be too high or it will operate too often on noise, which is undesirable because it increases the activity and hence reduces the TASI advantage. A compromise is needed to achieve minimum activity while maintaining satisfactory speech quality for even the weak talkers.

The potential impairment caused by the loss of speech energy below the sensitivity of the speech detector, and by the time delay required to set up a connection, can be made negligible by design. Another possible impairment is inherent in any line-concentration arrangement but it can be minimized to any desired degree by limiting the number of lines to be served. This factor is the possibility that the number of individuals who are talking or starting to talk at a particular instant will exceed the number of available channels. For example, telephone subscribers on a party line must share time with their neighbors on a single pair of wires leading to the central office. In addition, all trunk lines between two or more central offices may occasionally be busy and a call coming in at that instant will experience some delay. In a similar manner, TASI may occasionally be overloaded by an unusually large number of simultaneous talkers. In this case, however, the information cannot be delayed and any failure to make an almost instantaneous connection results in a loss of speech transmission, called a freezeout, until a channel becomes available. It appears that the effect on transmission quality is negligible as long as the percentage of lost speech, freezeout fraction, is less than 0.5%.

During periods of light traffic the system operates essentially the same as without TASI and hence there can be no TASI impairment. However, as the load increases, the amount of switching and possible impairment increases. The objective is to limit the number of busy lines to a point where the resulting impairment is not noticeable to the average listener and is not objectionable to the more critical listener.

The number of circuits that can be made available for service at the toll switchboard, for a freezeout fraction of 0.5%, is greater than the number of physical channels by the TASI multiplying factor shown in Fig. 3. As the number of channels in the group increases, the number of potential conversations approaches a limit set by the reciprocal of the average activity. It will be noted that a 36-channel group, such as is available in present submarine cable systems, achieves a substantial TASI advantage, but that the application of TASI to only 12 channels or less is much less attractive. For an average activity in the range of 35 to 40%, it appears that the use of TASI on a 36-channel group can at least double the number of available circuits and still maintain good quality.

The average number of simultaneous talkers, np, is the product of the number of busy lines, n, multiplied by the average activity p. When the average number of simultaneous talkers is approximately equal to the number of channels c, there is a 50% chance that a new talker will be frozen out momentarily. This does not mean that 50% of the speech is lost, but there is noticeable impairment. In order to allow for variations around the average and to achieve satisfactory quality, the average number of simultaneous talkers, np, should be limited to a value that will satisfy the approximate condition of $np+\sqrt{np}\leq c$. This approximate expression is based on a more exact derivation given in the Appendix. When

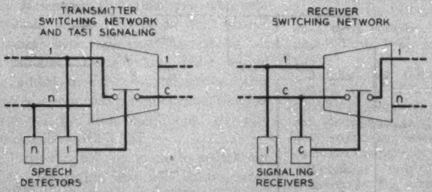


Fig. 2 (above). TASI block schematic

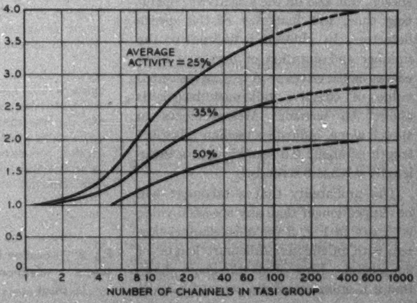


Fig. 3 (right). TASI advantage for a freezeout fraction of 0.5%

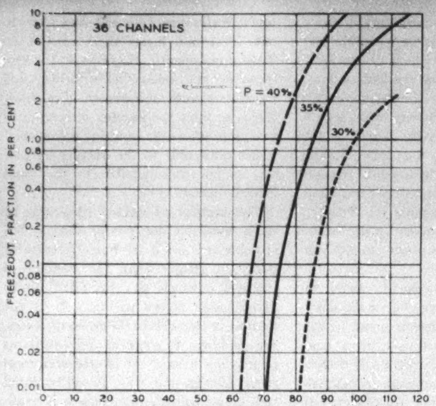


Fig. 4. Freezeout fraction versus activity and number of busy lines

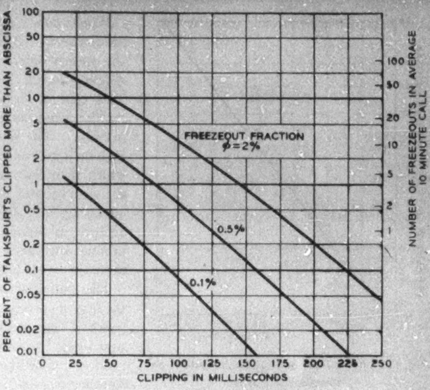


Fig. 5. Magnitude of clipping

this criterion is met, only a small percentage of the connections will experience any degree of freezeout except that required for the connect signaling code.

The expected freezeout fraction is shown in Fig. 4 as a function of the number of busy lines working on 36 physical channels for a talker activity of 30 to 40%. Listening tests have shown that a freezeout fraction of 0.1% is barely perceptible to a critical listener. A value of 0.5% is noticeable to the critical listener but is ordinarily not apparent to an average listener who is intent on understanding a message. A value of 2% would be somewhat objectionable, and a value of 5 to 10% would be definitely objectionable.

The average length of freezeout is roughly proportional to L/c, where L is the average talkspurt length and c is the number of channels. The average talkspurt length for TASI purposes depends not only on the characteristics of speech, but also on the sensitivity, operating time, and hangover of the speech detector. In addition, it depends on the number and duration of bursts of noise echoes on the circuit. A typical measured value of effective talkspurt length is about 0.6 seconds. Measurements of speech alone, indicate a median talkspurt length of about 0.6 seconds and a mean length of 1.5–1.8 seconds.[1]

The probability that a talkspurt will be clipped longer than any specified value is shown in Fig. 5 for a 36-channel system with an effective talkspurt length of 0.6 second. First of all, the chart shows that a 15-millisecond clip occurs for TASI signaling each time a new connection is needed. During the busy hour, this may occur on almost every talkspurt, but the number of operations will be reduced considerably in nonbusy hours. The three lines sloping toward the right represent three different freezeout fractions. For example, the middle curve labeled 0.5% indicates 0.5% of the talking time has been lost because a channel was not available. In this case, about 6% of the talkspurts are frozen out more than the 15-millisecond signaling time and 1% of the talkspurts are frozen out for 80 milliseconds or longer. A freezeout longer than 80 milliseconds can be expected about 4 times during an average 10-minute call. For comparison purposes, it is worth noting that the average syllable length is about 125 to 250 milliseconds. On the average, less than one talkspurt in 10,000 will be frozen out for as long as 1/4 second, and this will occur only once in about 30 calls during the busy hour.

As mentioned earlier, a talker holds a channel until his speech detector indicates that he has finished talking and until someone else needs a channel. This is true on all but three of the available channels. One channel is reserved full time for disconnect and other control purposes, and two talking channels taken at random are disconnected when possible in order to have a channel instantly available for the next talker.

Switching from channel to channel during a conversation also causes some variations in circuit noise and net loss. It is expected that a little noise will be introduced in the receiving circuits in order to prevent a completely dead receiver during the silent intervals. The variations in net loss are not expected to

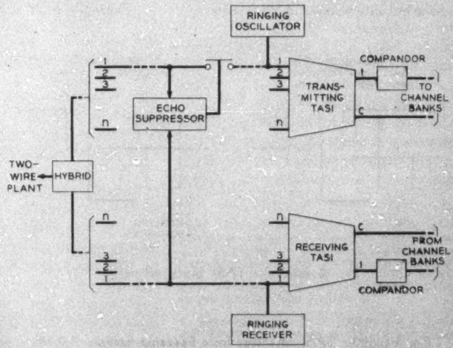


Fig. 6. Location in 4-wire telephone plant

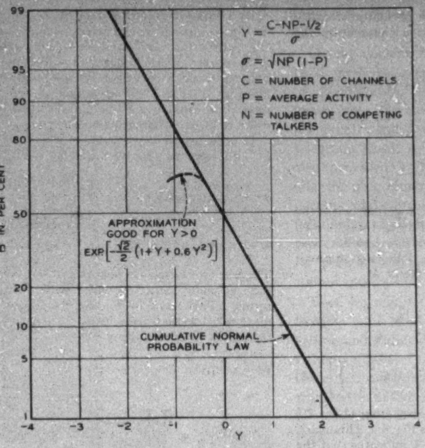


Fig. 7. Probability of freezeout $B_{(c,n,p)}$

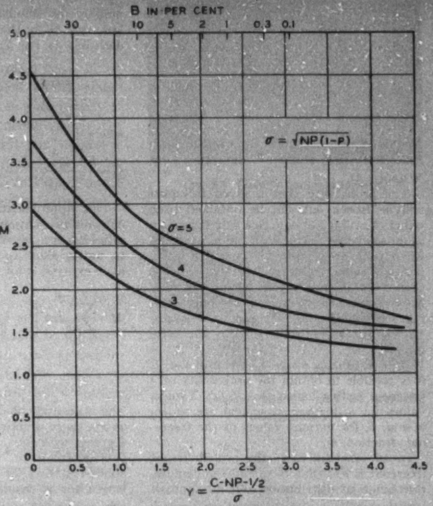


Fig. 8. Factor used in determining average length of freezeout

be important as long as the plant is maintained to meet the present objectives.

TASI is designed to take advantage of the characteristics of 2-way conversational speech. It requires a channel width of about 3,000 cycles in order to accommodate the tones needed for the rapid connect and disconnect signaling.

Telegraph or data systems that use frequency shift or frequency modulation will lock up the TASI speech detector 100% of the time and hence be transmitted satisfactorily. On the other hand, the use of on-off telegraph or data systems will require either a means for locking the speech detector or a separate channel outside the TASI complement. The high activity of telegraph and data systems requires an essentially full-time channel and hence reduces the number of conversations that TASI can yield

TASI will be located in the 4-wire part of the telephone plant, as shown in Fig. 6. It requires that split-type echo suppressors be used to insure that echoes cannot seize an outgoing circuit at either end. Compandors, which are to be added to deep sea submarine cable systems on a channel-by-channel basis as required by noise conditions, are also compatible with TASI.

Provision will be made to switch out TASI for testing purposes. In order not to affect service, circuits will be removed as they become idle until the number of circuits available to the switchboard is no greater than the number of physical channels before TASI is removed.

Because TASI is a time-sharing device, there are problems involved in transmitting supervisory and dialing pulses over TASI. TASI can work satisfactorily with the present ringdown manual basis but it is obvious that continuous supervision by means of a steady tone during the idle time cannot be used. Likewise, dial pulses cannot compete for a TASI channel on the same basis as a talker because TASI clipping would cause signaling errors. Various schemes using spurt signaling are being investigated for use when dialing over TASI is required.

Appendix

The TASI principle has been recognized for many years, and several individuals, including R. I. Wilkinson, have contributed to the mathematical analysis of speech interpolation. The following summary is based on a more complete analysis given in an unpublished 1953 memorandum by H. Cravis.

Assume that there are n independent talkers, each of whom has an average activity p competing with an individual test call (total number of talkers $= n+1$). At any particular instant, the probability that the number of simultaneous talkers will equal or exceed c (where c is less than n) is given by the cumulative binomial distribution, $B_{(c,n,p)}$.

$$B_{(c,n,p)} = \sum_{x=c}^{n} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \qquad (1)$$

The probability of a freezeout lasting longer than t seconds is given by $B_{(c,n,\theta)}$ where $\theta = pe^{-t/L}$ for an exponential talkspurt distribution, or $\theta = p(1-t/L)$ for a constant talkspurt length. The assumption of an exponential distribution of talkspurt lengths is the best simple approximation to the experimental data. The use of a constant talkspurt length does not change the over-all result significantly. In this expression L is the average talkspurt length.

It can be shown that the average freezeout length for an exponential talkspurt distribution is

$$t_F = L \sum_{k=c}^{n} \frac{1}{k} \frac{B_{(k,n,p)}}{B_{(c,n,p)}} \qquad (2)$$

$$= \frac{L}{c} M \qquad (3)$$

$$\text{where } M = \left[1 + \frac{c}{c+1} \frac{B_{(c+1,n,p)}}{B_{(c,n,p)}} + \frac{c}{c+2} \frac{B_{(c+2,n,p)}}{B_{(c,n,p)}} + \ldots \right] \qquad (4)$$

This series converges rapidly so only a a few terms are needed for sufficiently accurate results.

The fraction of the speech that is frozen out is

$$\Phi = \frac{t_F B_{(c,n,p)}}{L} = \frac{M}{c} B_{(c,n,p)} = \sum_{k=c}^{n} \frac{1}{k} B_{(k,n,p)} \qquad (5)$$

Tables of the binomial distribution are available for accurate computations. However, in the range of interest to TASI, this function is almost identical with the cumulative normal probability law, that is,

$$B_{(c,n,p)} \approx \frac{1}{\sqrt{2\pi}} \int_{u=y}^{\infty} exp\left[\frac{-u^2}{2}\right] du \qquad (6)$$

where

$$y = \frac{c - np - 1/2}{\sqrt{np(1-p)}}$$

In this expression the mean value is $np + 1/2$ and the standard deviation $\sigma = \sqrt{np(1-p)}$.

The probability that a given talkspurt will be frozen out can be obtained from either equation 1 or equation 6 and is shown as a solid line on Fig. 7. For example, when $y = 1$, 16% of the talkspurts will experience some degree of freezeout because there are more talkers than channels. Most of these freezeouts will last only a few milliseconds so the percentage of speech that is lost is less than 16% by the ratio t_P/L.

By substituting $\theta = pe^{-t/L}$ for p in Fig. 7, it is possible to obtain the probability of a freezeout lasting longer than t/L. Typical values for a 36-channel system are shown in Fig. 5, for various values of the freezeout fraction Φ.

In all practical cases the probability of freezeout is less than 50% and the cumulative binomial distribution and its approximation, the cumulative normal-probability law, can be further approximated by the dash line on Fig. 7, that is,

$$B_{(k,n,p)} \approx exp\left[-\frac{\sqrt{2}}{2}(1 + y_k + 0.6 y_k^2)\right] \quad \text{for } B < 0.5 \qquad (7)$$

where

$$y_k = \frac{k - np - 1/2}{\sigma} = y_c + \frac{k - c}{\sigma}$$

This approximation, together with the omission in equation 4 of the factors $c/(c+1)$, $c/(c+2)$, etc., which are less than but close to unity, leads to the following approximate expression for the factor M

$$M \approx \sum_{k=c}^{\infty} exp\left[-\frac{\sqrt{2}}{2}\frac{k-c}{\sigma} \times \left(1 + 1.2 y_c + 0.6\frac{k-c}{\sigma}\right)\right] \qquad (8)$$

The magnitude of M varies from about 1.5 to 3 or more as shown in Fig. 8. This chart, together with equation 3, indicates that the average freezeout, when it occurs, is only about 1/10 to 1/20 of the average talkspurt length in a 36-channel system.

By the use of equation 5 and Figs. 7 and 8, Fig. 4 can be constructed and similar charts for other values of c can be obtained. A cross plot of such data for a constant value of Φ yields curves of TASI advantage like Fig. 3. The TASI advantage is $n + 1/c$ since there must be $n + 1$ potential talkers in order to have n competing talkers.

In order to achieve a value of Φ less than 1%, it is necessary for $y > 1$, that is,

$$y = \frac{c - np - 1/2}{\sqrt{np(1-p)}} > 1$$

or $c > np + \sqrt{np} - \sqrt{np}(1 - \sqrt{1-p}) + 1/2$

The last two terms are opposite in sign and each is small compared with the first two terms. This leads to the simple approximate relation that a practical TASI should be limited to the following condition:

$$np + \sqrt{np} < c$$

Reference

1. CHARACTERISTIC TIME INTERVALS IN TELEPHONE CONVERSATION, **A. C. Norwine, O. J. Murphy.** *Bell System Technical Journal*, New York, N. Y., Apr. 1938, pp. 281–91.

A Transistorized 20-Channel Carrier Telegraph Terminal

T. M. GRYBOWSKI
NONMEMBER AIEE

W. G. VIETH
MEMBER AIEE

SINCE ITS INTRODUCTION in 1938 the FM (frequency-modulated) carrier has found ever-increasing use in the Western Union Telegraph Company. The gradual evolution from the type-*15* to the current type-*40* carrier telegraph terminal has shown an increase from two telegraph channels per bay, to 10 channels in the same space. The equipment herein described mounts two 10-channel groups per bay along with associated power supplies, bias and leg controls, and metering and monitoring facilities. This design has been designated as the type-*60* channel terminal.

Physical Description

The design of the type-*60* channel terminal decided upon in the planning phase of the project included consideration of the degree of size reduction made possible in a completely transistorized equipment, the practicability of manufacture, and the ease of installation and maintenance. These considerations resulted in the complete equipment being mounted on what is now known as channel terminal bay *601-A*.

The *601-A* bay consists of a standard 19-inch equipment rack upon which the equipment is mounted and occupies 9 feet 4 inches of vertical rack space. The components of the equipment are mounted in order from top to bottom as shown in Fig. 1. At the top are the terminal blocks where the external carrier and d-c signal circuits are terminated and where the various components on the bay are interconnected. Next is the carrier coupler panel, not visible in Fig. 1, that contains four line-isolating transformers, one for each direction of transmission for two channel groups. The power distribution panel contains d-c power-indicator-type fuses that distribute leg and transistor battery to the individual channels. This location, near the top of the bay, provides good visibility of the blown fuse indicator lamps without occupying more useful operating space. The power is brought to this panel from the potential cabinet at the bottom of the bay by a flexible conduit in the space between the rack channels.

Distribution of positive and negative 125-volt leg battery and negative 25-volt transistor battery to each channel takes place through the wiring duct shown at the left side of the rack. Below this panel are five rack shelves, each mounting two of the 10 channels of channel group 1. Beneath channel group 1 and located at a convenient height are grouped the metering, monitoring, and control facilities for all channels on the bay. The 10 channels of group 2 and a potential cabinet and power entrance box complete the bay. The potential cabinet contains the leg battery and transistor battery rectifiers, the main fuses, power switch, and a utility outlet. The power entrance box provides termination for external 115-volt a-c power, and circuit and frame grounds.

A single type-*60* channel terminal consists of a plug-in transceiver *650-A* and

---

Paper 59-85, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 23, 1958; made available for printing December 17, 1958.

T. M. GRYBOWSKI and W. G. VIETH are with the Western Union Telegraph Company, New York, N. Y.

Vol.
Pt. 1
1959

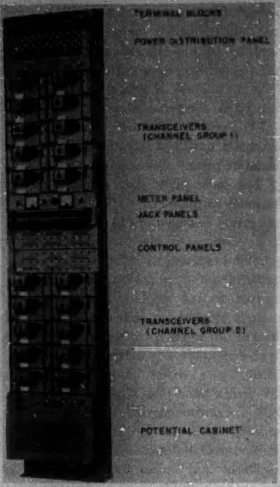


Fig. 1. Carrier bay 601-A

associated tuned networks as shown in Figs. 2 and 3, and remotely located leg current and receiving bias controls. The transceiver plugs into its supporting shelf, illustrated in Fig. 4, and is secured by means of a camlock fitted with a shaft which terminates on the front panel in a 2-position OPEN and LOCK knob. The networks plug into the transceiver chassis and are fastened in place by screws. Each transceiver contains electronic sending and receiving circuits which are identical for all channels. The operating frequency of the transceiver is determined by the particular tuned networks associated with it.

Circuit considerations led to the use of a receiving polar relay type *202-A* which mounts in a socket on the transceiver face. To meet the wide variety of d-c circuit conditions encountered in the telegraph plant, a "stunt box" socket is also provided. The stunt box or leg adaptor plug illustrated in Figs. 2 and 3 permits make-break leg operation on a full-duplex, half-duplex, or duplex repeater basis. Other plugs are available for polar, telex, and a wide variety of special applications.

Each leg adaptor plug is secured by a locking latch that can be released by raising and sliding its handle to the right. Other controls on the transceiver face include a transmitter ON–OFF switch, a transmitter LEVEL control, and a number of test-pin jacks. Only sufficient test points are provided to determine if the unit is defective so it may be replaced immediately with a spare unit to maintain service. Fig. 5 shows a rear view of the transceiver with the networks removed and the hinged side panels open to expose the transistorized subpanels.

The control facilities for the bay shown in Fig. 1 consist of a jack panel and control panel for each channel group, and a common meter panel. The jack panel permits the meter panel to be corded into the sending and receiving leg circuits of any particular channel so that testing and regulating functions may be performed. The control panels contain a sending and receiving leg rheostat and a receiving bias control for each channel. Practically all of the heat dissipation in the bay is concentrated in the leg rheostats.

Fig. 2. Channel terminal type 60

Fig. 3. Channel terminal type 60, partially disassembled

This was a major consideration in the rack layout to avoid exposure of transistor circuitry to excessive temperatures. This resulted in an open framework mounting for the rheostats and a louvered front cover to provide maximum ventilation.

The back of the rack shelves visible in Fig. 4 are also perforated for ventilation of the battery tap resistors mounted on the rear of the transceiver chassis. The perforated and louvered covers on the potential cabinet and control panels, in conjunction with the space between the rack channels, create a chimney effect causing air flow through the bay. This prevents excessive temperature rise. Measurements indicate a maximum rise within any transceiver of only 12 C

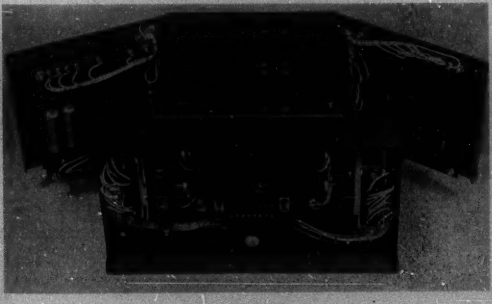


Fig. 4 (left). Close-up of bay with one channel terminal removed from shelf to show shelf detail

Fig. 5 (above). Transceiver 650-A, rear view with hinged side panels open

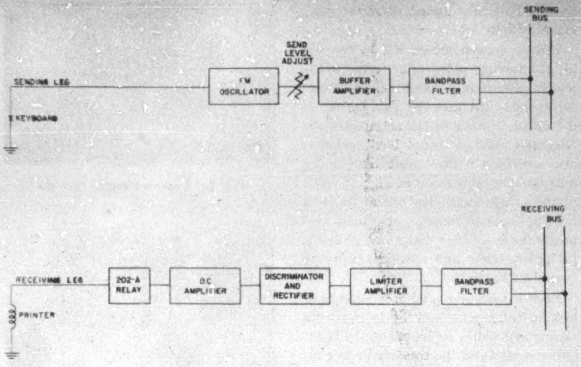

SENDING BUS
SEND LEVEL ADJUST
SENDING LEG
FM OSCILLATOR
BUFFER AMPLIFIER
BANDPASS FILTER
KEYBOARD
RECEIVING BUS
RECEIVING LEG
202-A RELAY
D.C. AMPLIFIER
DISCRIMINATOR AND RECTIFIER
LIMITER AMPLIFIER
BANDPASS FILTER
PRINTER

Fig. 6. Channel terminal type 60, block diagram

(degrees centigrade) above ambient room temperature.

The *601-A* bay represents a marked change in design of carrier equipment and methods of mounting. Panels and mounting shelves are now essentially a part of the rack, and the plug-in shelf-mounting concept for the transceivers and plug-in power supplies make for ease of shipment, installation, inspection, maintenance, or replacement. The bay design permits assembly of the permanently mounted components and complete wiring of the bay to be performed at the factory. At installation it is only required to set the bay in place, wire the external transmission facilities to the terminal blocks at the top of the bay, and a-c power and grounds to the power distribution box. The transceivers and power supplies are then plugged in place to complete the installation. The plug-in components comprise 340 pounds of the 650-pound total installed weight. The *601-A* bay required 60% less office space and 65% less power per channel than currently used equipment.

Circuit Description

The type-60 channel terminal is fully compatible with the types 20, 30, and 40 currently in use throughout the telegraph plant. In each type-60 channel terminal the marking and spacing signals frequency-modulate an audio-frequency carrier which is then combined with other channel frequencies for transmission over a carrier subband facility. The operating speed of these circuits is 75 bauds to provide suitable margin for teleprinters operating at 100 words per minute. Frequency modulation of the transmitted carrier is accomplished by lowering the carrier frequency 35 cycles below channel mid-frequency for a marking pulse, and by raising the frequency 35 cycles above mid-frequency for a spacing pulse. Ten individual channels are provided, spaced at 150-cycle intervals from 375 to 1,725 cycles per second. Both channel groups 1 and 2 of bay *601-A* utilize the same frequency allocations.

The basic block diagram of a carrier channel as arranged for make-break leg operation is shown in Fig. 6. The closing and opening of the sending leg by a printer keyboard or other keying source causes the frequency of an oscillator to be shifted to the marking or spacing frequency by a capacitive reactance switched into or out of circuit by means of a diode bridge switch. The carrier signal is amplified by a buffer stage and applied through a band-pass filter to the sending bus. Continuous adjustment of the sending level is accomplished with a gain-control potentiometer coupling the oscillator and buffer amplifier. At the receiver, a suitable band-pass filter selects the desired carrier signals from the receiving bus. These signals are amplified and limited and applied to a series-tuned discriminator. The resulting d-c polar telegraph signals pass through a d-c amplifier and operate a type-*202-A* polar relay.

Transmitter Circuit

Since the general theory of FM telegraphy has been adequately covered[1,2] this paper will describe how the new terminal differs from existing vacuum tube types particularly with respect to design features necessitated by the use of transistors.

The basic circuit of the transmitter arranged for neutral leg operation is shown in Fig. 7. Functionally, the trans-

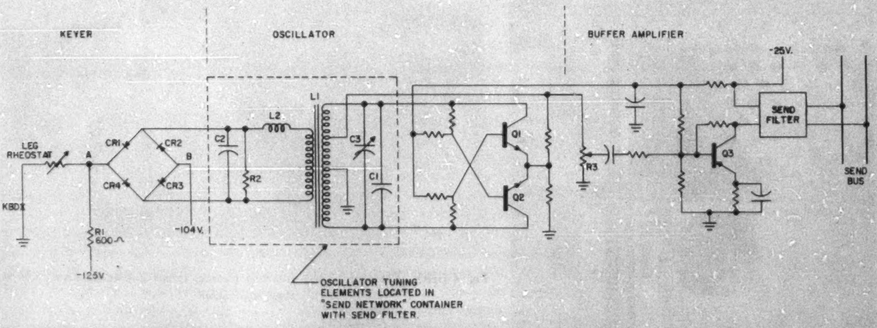

KEYER
OSCILLATOR
BUFFER AMPLIFIER
-25V.
LEG RHEOSTAT
A
B
CR1
CR2
CR3
CR4
KBD
R1 600-Ω
-125V
-104V.
C2
L2
R2
L1
C3
C1
Q1
Q2
R3
Q3
SEND FILTER
SEND BUS
OSCILLATOR TUNING ELEMENTS LOCATED IN "SEND NETWORK" CONTAINER WITH SEND FILTER

Fig. 7. Transmitter circuit

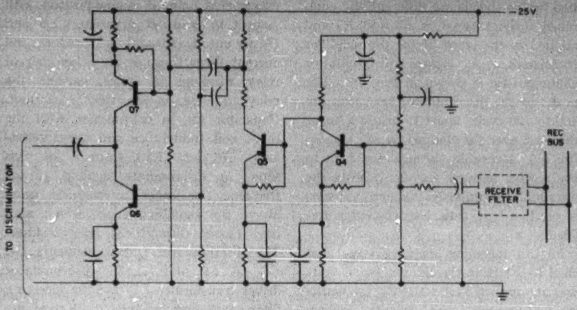


Fig. 8. Receiver limiter-amplifier

mitter consists of an oscillator, keyer, and buffer amplifier. When the sending leg is closed and the leg rheostat adjusted for a current of 70 ma (milliamperes), a potential of approximately minus 83 volts exists at point *A*. Since point *B* is held at minus 104 volts by a voltage divider, the anodes of diodes *CR*1 through *CR*4 are more negative than the cathodes by 21 volts, causing them to be nonconducting.

The oscillator frequency is determined by *L*1, *L*2, *C*1, *C*2, *C*3, and in small part by the emitter base or diffusion capacitance of transistors *Q*1 and *Q*2. When the sending leg is open, point *B* becomes positive with respect to *A*, and the diodes conduct. As a result, capacitor *C*2 is short-circuited and the resonant frequency is determined by *L*1, *L*2, *C*1, and *C*3. By design, the two resonant frequencies differ by 70 cycles. The use of the diode bridge configuration minimizes the induction of a keying transient into the oscillator tuned circuit.

The oscillator circuit is of the push-pull, common (grounded) emitter type similar in operation to the vacuum tube equivalent. The similarity is obvious if the correspondence between collector, base, and emitter of the transistor and the plate, grid, and cathode, respectively, of the vacuum tube is considered. N-p-n transistors are used because they provide greater frequency stability with temperature change than p-n-p types. This is true since the emitter-base capacitance of the typical n-p-n is less than one tenth that of the typical p-n-p. The corresponding change of capacitance due to temperature change is reduced proportionately. An output is taken from a tap on the primary of *L*1 and applied through the send level control *R*3 to the base of buffer amplifier *Q*3. Potentiometer *R*3 permits a continuous adjustment of the transmitter carrier level. A carrier level of plus 7.5 dbm (power in decibels referred to 1 milliwatt) is available at the output of the sending filter. The input impedance of the send filter is 7,500 ohms and the output impedance 600 ohms. The sending filter and the oscillator circuit components *C*1, *C*2, *C*3, *L*1, *L*2, and *R*2 are mounted in the sending network located on the left side of the transceiver and visible in Fig. 1 through 4. The network connectors are polarized to prevent the insertion of a sending network into a receiving network position or vice versa. A sending network is identified by the letter designation of the channel on the handle as well as by a type number and channel frequency stenciled on the top of the network container, visible in Fig. 3. The hole in the handle allows access to trimmer *C*3, which provides frequency adjustment.

RECEIVER CIRCUIT

The receiver circuit is diagrammed in Figs. 8 and 9. The received carrier signal passed by the receiving filter enters the limiter-amplifier, consisting of transistors *Q*4 through *Q*7, see Fig. 8. Here amplitude variations in the incoming signal are removed and the signal is raised to the constant level required to drive the discriminator. The discrimi-

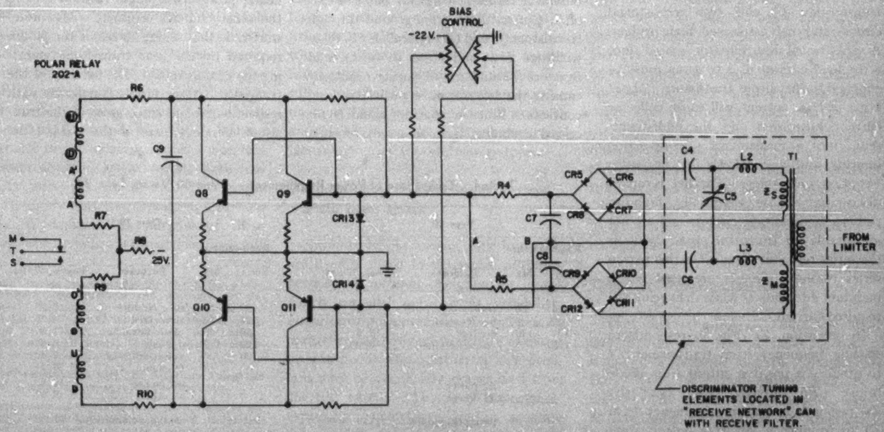


Fig. 9. Receiver discriminator and d-c amplifier circuits

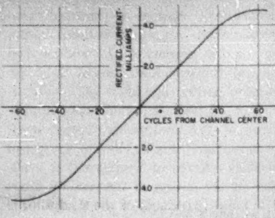


Fig. 10. Typical discriminator characteristic

nator driver transistors, $Q6$ and $Q7$, are connected in a complementary symmetry arrangement to provide uniform loading of the discriminator transformer $T1$ and also to simplify the circuit. For signal levels of less than −45 dbm, $Q4$ acts essentially as a class A amplifier. For larger signals limiting begins. Transistor $Q5$, directly coupled to $Q4$, further amplifies and limits the signals. For inputs in excess of −55 dbm the gain of $Q4$ and $Q5$ is sufficient to saturate the output transistors, resulting in an 18-volt peak-to-peak square wave at primary of $T1$.

The discriminator converts the incoming FM carrier to d-c polar signals. These signals are in turn amplified by transistors $Q8$ through $Q11$ to operate a polar relay. Four transistors are used as a protective measure since, under the 3-fold circumstance of loss of incoming carrier, an ambient temperature of 55 C, and proper or improper setting of the receiver bias control, the dissipation ratings of the transistors can be exceeded.

The discriminator consists of driver transformer $T1$ and two series-tuned circuits and their associated diode bridges. A series-tuned discriminator is used since a d-c load current of 3 to 4 ma must be supplied to the output transistors. About 1 ma of base current will cause collector current saturation and the added current gives a measure of post-detection limiting and also allows for transistor variation and aging. Parallel resonant discriminators are essentially high-impedance high-voltage circuits and are not as suitable for transistor applications as the series resonant type. For this reason series resonant circuits are used which maintain adequate Q while delivering the required current.

In operation, Zs (Fig. 9) responds to spacing frequency from transformer $T1$ to produce a positive output from diode bridge $CR5$ through $CR8$. Similarly, Zm responds to marking frequency from $T1$ to produce a negative output from diode bridge $CR9$ through $CR12$. The two bridge signals are differentially combined through resistors $R4$ and $R5$ and applied to the input of the d-c amplifier transistors. A typical discriminator characteristic is shown in Fig. 10. The peaks in the discriminator responses are 100 to 120 cycles apart to insure a linear response over the channel band. Capacitor $C5$, accessible through the hole in the receive network handle, adjusts the discriminator tuning so that zero response is produced at the channel center frequency.

For an incoming marking signal applied to $T1$ from the limiter, equal voltages are applied to Zm and Zs, the marking and spacing tank circuits, respectively. Zm, tuned close to the marking frequency, provides a low impedance to the induced voltage; Zs, tuned away from the marking frequency, presents a high impedance. As a result, more current flows through the marking bridge, $CR9$ through $CR12$, than through the spacing bridge, $CR5$ through $CR8$. Capacitors $C7$ and $C8$ act to filter the rectified output of the diode bridges and remove the harmonics of the detected carrier frequency. Because a larger voltage is developed across $C8$, point A becomes negative with respect to point B. Current flows from point B through diode $CR14$ and the emitter-base circuits of transistors $Q8$ and $Q9$ to cause these transistors to conduct. The voltage drop across diode $CR14$ biases the bases of transistors $Q10$ and $Q11$ to cutoff.

A negative voltage on the base of a p-n-p transistor results in conduction; conversely a positive potential at the base will result in collector current cutoff.[3] Current flows from ground through transistors $Q8$ and $Q9$ in parallel, $R6$, relay windings U-D and A'-A in series, and resistors $R7$ and $R8$ to negative battery causing the relay to close on its marking contacts. When a spacing signal is received a similar train of events occurs, except that point A is now positive with respect to point B, transistors $Q8$ and $Q9$ are cutoff, $Q10$ and $Q11$ conduct, and current flows through the other pair of relay windings in series, causing the relay to close on the spacing contact. Capacitor $C9$, in conjunction with the relay coil inductance and series resistances $R6$ and $R10$, form a low-pass filter to discriminate against carrier frequency or other interference lying above the signalling range which may have passed through the receiver filter. The discriminator tuning elements $C4$, $C5$, $C6$, $L2$, and $L3$, the discriminator driver transformer $T1$, and the receiving filter are mounted in the plug-in receiving network located on the right side of the transceiver as shown in Figs. 1 through 4. The receiving network is distinguishable in that it mounts a designation clip and omits the channel letter designation on the handle. As with the sending network, the network type number and channel frequency are also stenciled on the top of the container.

Power Consumption

The extent of miniaturization and reduction in power consumption achieved with this equipment was possible only through the use of transistors. Their low-voltage low-current requirements, without the need for filament-supply voltage, caused a significant reduction in power consumption. Each transceiver, containing 11 transistors, requires only 1.25 watts of power for its operation (25 volts at 50 ma) exclusive of telegraph leg power. This is less than the heater power for a single vacuum tube in the older channel terminal. Also noteworthy is that about 30 ma of the 50 are required by the four transistors operating the output relay. The polarity of the transistor battery supply is negative with respect to ground since most of the transistors employed are of the p-n-p type.

Table I. Comparison of Power Requirements: Type 60 Versus Type 30

Type 60				Type 30			
Requirement				Requirement			
Volts	Ma	Purpose	Power, Watts	Volts	Ma	Purpose	Power, Watts
−25	50	transistors	1.25	+120	18	tube circuits	2.16
−22	5	bias control	0.11	−120	73	tube circuits	8.76
−125	140	d-c legs	17.5	50	20	bias control	1.0
−125	7	voltage divider	0.875	−120	140	d-c legs	16.8
				−120	20	voltage divider	2.4
						filament	26.2*
		leg power	17.5			leg power	16.8
		all other	3.8			all other	40.5
		total per channel	21.3			total per channel	57.3

* 25 volts a-c, 1.05 amperes.

A comparison of the power requirements of the type *60* and the type *30* is shown in Table I. Of interest is the fact that a bay of 20 type-*60* channels requires less total power than one of 8 type-30 and occupies no greater floor space.

MAINTENANCE

In order to reduce the need for and facilitate maintenance where required, a certain design philosophy has been followed. These ideas may be grouped generally under three headings: quality components, use of test points and other failure indicating devices, and plug-in construction. The quality components employed include hermetically sealed paper and solid electrolyte tantalum electrolytic capacitors for coupling and bypass purposes in the transistor circuits. Silvered mica and Mylar film capacitors are used in the tuned circuits of the sending and receiving filters, oscillator, and discriminator tanks. Ferrite cup-core inductors are used in the oscillator and discriminator tuned circuits. Industrial or computer quality transistors are used to insure long life. Silicon junction rectifiers are employed in the power supplies.

Test points, visible in Figs. 1 and 2, on the front of the channel transceiver, facilitate testing and aid in localizing faults. The indicator fuses located on the power distribution panel and the potential cabinet signal the failure of a particular channel or power supply. Primary power and leg circuit fuse indicators are of the neon glow lamp type. Transistor and bias circuit fuse indicators are of the incandescent lamp variety.

Conclusion

The design objectives of this development have been substantially achieved. The advantages of space saving, flexibility and self containment as to power supply and test facilities, and reduced heat dissipation, enhance the value of this equipment. Both small offices where load requirements make but a single voice-frequency band necessary or large installations where space problems, floor loading, and air-conditioning load are becoming serious are likely locations for the use of this equipment. The use of reliable and long-lived components should considerably reduce the cost of maintenance.

References

1. Frequency-Modulated Carrier Telegraph System, F. B. Bramhall, J. E. Boughtwood. *AIEE Transactions*, vol. 61, Jan. 1942, pp. 36–39.

2. An FM Telegraph Terminal Without Relays, F. H. Cusack, A. E. Michon. *Western Union Technical Review*, New York, N. Y., Oct. 1947.

3. Transistor Characteristics and Circuits, H. C. Likel. *Ibid.*, Oct. 1957.

Tests Conducted Over Highly Reflective Terrain at 4,000, 6,000, and 11,000 Megacycles

A. OXEHUFWUD
NONMEMBER AIEE

DURING the past 10 years the American Telephone and Telegraph Company has conducted tests at 4,000 mc (megacycles) over most major microwave radio relay routes prior to the initial installations. As the necessity for further expansion arises, it is planned to add 6,000-mc and, possibly, 11,000-mc circuits to the existing routes. For this reason antenna height versus path loss tests were performed over four paths with highly reflective terrain in Colorado and Wyoming. The purpose of these tests was to collect information to permit the evaluation of performance at the higher frequencies on existing paths from the results of earlier 4,000-mc tests when a common antenna is used for all frequencies. To obtain comparable results simultaneous measurements were made at 3,950 and 6,310 as well as 3,950 and 11,200 mc. The 3,950-mc measurements were used for reference purposes.

Purpose of Antenna Height Versus Path Loss Tests

Antenna height versus path loss tests are usually made to determine optimum permanent antenna heights for paths which require high circuit reliability. Locations and elevations of obstructions and reflecting surfaces which are of major importance in the solution of this problem may be accurately determined from the test results. This is of importance, especially when there are no accurate topographic maps available. When only one reflecting surface exists in a path, the problem is usually a simple matter. More frequently, however, there are multiple surfaces. The reflections from each surface may not be severe but, if even numbered Fresnel zone reflections from two or more of these surfaces should happen to coincide, cancellation of the desired signal can be intolerable. Therefore it is very important that the location and elevation of each reflecting surface be known and, also, the severity of the reflections. When this has been established through antenna height versus path loss tests, it is usually possible to select antenna heights which will prevent coinciding reflections regardless of changes in bending of the rays caused by variations in atmospheric refraction. In 1950 and 1955, 4,000-mc tests were made over the paths which are to be discussed in this paper. Thus the major obstructions and reflecting surfaces were already known. It is interesting to note that the tests made in July and August of 1950 and those made in April and early May of 1957, checked almost exactly, reading for reading.

Types of Test Equipment

The test equipment employed to measure the path loss at 3,950 mc consisted of a Western Electric Company, Inc., type *TE-2* transmitter and receiver. The 6,000-mc equipment was RCA (Radio Corporation of America) *TVM-1A*, and the 11,000-mc equipment Lambda-Pacific Engineering, Inc., *1200K*.

Paper 59-86, recommended by the AIEE Radio Communication Systems and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 14, 1958; made available for printing December 11, 1958.

A. Oxehufwud is with the American Telephone and Telegraph Company, Kansas City, Mo.

Appreciation is expressed to J. H. Jacquier of the Bell Telephone Laboratories and R. L. Corbridge, E. J. Runner, and others in the American Telephone and Telegraph Company who participated in these tests.

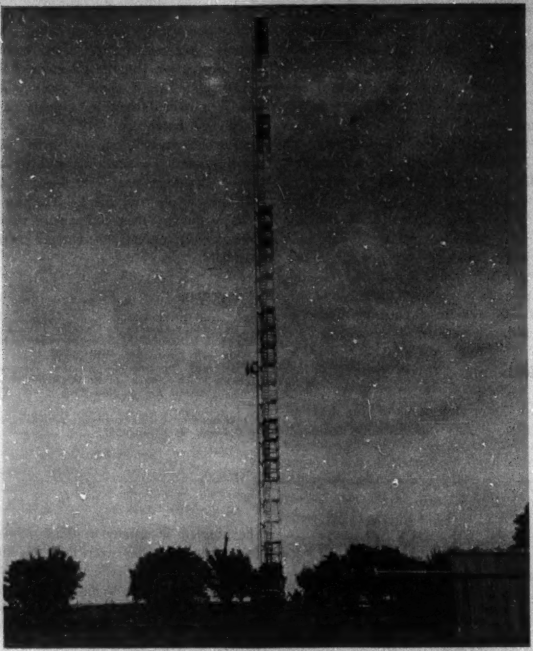


Fig. 1. 200-foot portable tower

With the 4,000-mc equipment 57-inch parabolic antennas were used. The 6,000- and 11,000-mc. equipments were equipped with 48-inch parabolic antennas. Vertical polarization was used with the 4,000-mc, horizontal with the 11,000-mc and both were tried with the 6,000-mc tests. To obtain the expected level at the receiver input for free space transmission the transmitter output is measured and added to the gains of the two antennas. This sum is then subtracted from the computed free space loss between two isotropic antennas for the path length between the transmitter and the receiver. A test signal is then applied to the receiver input at the free space level just mentioned. An attenuator inserted ahead of the main intermediate-frequency amplifier is adjusted to give a standard reading on a meter which measures the rectified output from the amplifier. The attenuator setting is recorded and used as a reference. Similarly, during the tests the attenuator setting required to give the standard reading on the meter for the received signal is recorded and, using the reference setting, the actual path loss is calculated. Portable towers 200 feet in height were equipped with tracks and two carriages mounted on these tracks, each holding one radio-frequency equipment with its associated antenna. The carriages were mounted in tandem one above the other and spaced about 5 feet between the center line of the antennas.

Fig. 1 shows a 200-foot portable tower. Fig. 2 shows the vehicle housing the receiver control units and the calibration equipment. The radio-frequency unit with the antenna is hoisted up and down the tower by means of a hand winch, which is equipped with an indicator giving the height of the center line of the antenna above ground. Fig. 3 shows the method for mounting the 4,000-mc equipment on the carriage below the 6,000-mc RCA equipment. During the 11,000-mc tests the Lambda Pacific equipment replaced the RCA unit.

Testing Methods

Antenna height versus path loss tests consist of path loss measurements at various antenna height combinations. For instance, the antennas at the transmitter and the receiver are both raised simultaneously in equal steps, or one antenna is held fixed at a certain height while the other is moved. Usually moving an antenna in 5-foot steps and measuring the path loss is satisfactory but, if the terrain is highly reflective, even a few inches at a time can be too much to notice the effect of every reflection. As the frequency is increased, the spacings between the Fresnel zone effects on the desired signal become smaller necessitating that the antennas be moved cautiously in very small steps. To analyze the test results a curve is plotted showing the path loss versus the antenna heights.

Because of irregularities in the carriage tracks and backlash in the antenna mountings it is always necessary to orient the antennas both vertically and horizontally for maximum received signal after they have been moved to a new position. This is done by remotely controlled motors. When no significant reflections are encountered or when the angle between the direct and reflected signals is small, this procedure is satisfactory. On the other hand, with a highly reflective surface near one end of a path with an elevation considerably lower than at the distant end, the problem becomes more involved. The angle between the direct and the reflected signals may then be so large that the directivity of a high-gain antenna partially differentiates between the two signals. Then the maximum or minimum signal is not received when the antenna at the end nearest to the reflecting surface is oriented in the direction of the direct signal path.

This may best be explained by the following example in which the angle between the direct and the reflected rays is large and the antennas are set at such heights that the path of the reflected signal is an odd multiple of one-half wavelength longer than the direct path (an odd-numbered Fresnel zone). If the vertical angle of the antenna nearest to the reflecting surface is changed from one extreme to the other, the received signal will increase gradually and go through a maximum, when the combined effect of the direct and reflected rays is the greatest. This will occur when the antenna is pointed below the direction from which the direct ray is arriving. On the other hand, should the antennas be set at such heights that the path of the reflected signal is a multiple of one wavelength longer than the direct path (an even-numbered Fresnel zone), the received signal will behave differently. If the antenna nearest to the reflecting surface is

Vol.
Pt. 1
1959

Fig. 2 (above). Vehicle housing receiver control units

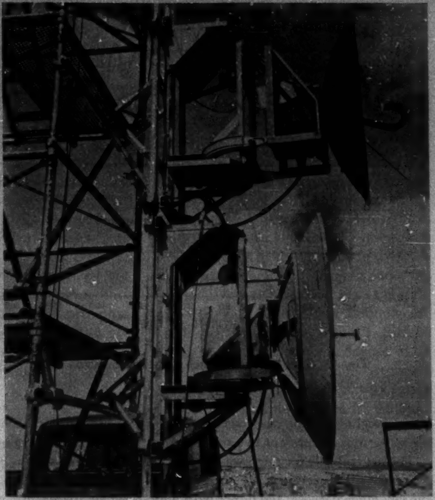


Fig. 3 (right). Mounting arrangement for 6,000-mc equipment (top) and 4,000-mc equipment (bottom)

pointed several degrees up from the horizontal and then is slowly oriented downward, the cancellation effect between the signals will vary as follows. As the antenna is oriented more toward the direct beam, the signal increases until a maximum is obtained. This occurs before the antenna is pointed in line with the path of the direct beam because the reflected signal is beginning to interfere. Thereafter, the received signal decreases and a minimum is obtained when maximum cancellation occurs between the two signals. As the antenna is pointed further downward, the signal again reaches a maximum when the reflected signal is stronger than the direct signal because of the antenna discrimination pattern. This will occur when the antenna points in a direction which is lower than the path of the reflected beam. During tests over this type of terrain, special attention is paid to this problem as it can be of importance in determining the proper orientation of the permanent antennas, especially when a common antenna is used for several frequency bands.

Paths Tested

Tests using 6,310 and 11,200 mc were made over four paths. Two of them were selected for the great difference in elevation between the sites and also for the smooth reflecting surfaces near the low end of each path. These were Buckhorn Mountain to Greeley in Colorado, and Rawlins to Creston in Wyoming. The other two paths, Bitter Creek to Creston, and Evanston to Church Butte, both in Wyoming, were chosen because they contained well-defined multiple reflecting surfaces more toward the center of the paths, so that the angle between the direct and reflected paths was small with respect to the beam widths of the antennas.

The characteristics of these paths and the antenna heights during the tests are as follows:

Buckhorn Mountain to Greeley:

Path length.............. 44.1 miles
Elevation at Buckhorn Mountain..............8,300.0 feet
Elevation at Greeley.......4,860.0 feet
Transmitter antenna at Buckhorn Mountain fixed at................ 52.5 feet

Receiver antenna at Greeley varied. Reflecting surfaces consisting of bare cultivated fields located within 4 miles of Greeley site

Rawlins to Creston:

Path length.............. 27.7 miles
Elevation at Rawlins......7,712.0 feet
Elevation at Creston......7,182.0 feet
Transmitter antenna at Rawlins fixed at........ 17.0 feet

Receiver antenna at Creston varied. Reflecting surfaces consisting of grass-covered desert located within 5 miles of Creston site.

Bitter Creek to Creston:

Path length.............. 39.3 miles
Elevation at Bitter Creek..7,096.0 feet
Elevation at Creston......7,182.0 feet

Transmitter antenna at Bitter Creek and receiver antenna at Creston both varied. Reflecting surfaces consisting of grass-covered desert located immediately in front of Bitter Creek and 6 to 7 miles from Creston

Evanston to Church Butte:

Path length.............. 36.8 miles
Elevation at Evanston....7,686.0 feet
Elevation at Church Butte.7,055.0 feet

Transmitter antenna at Evanston and receiver antenna at Church Butte both varied. Reflecting surfaces consisting of grass-covered desert and sparse sage brush close to both sites and 12.5 miles from Evanston.

Results of Tests

The tests confirm that the 6,000- and 11,000-mc transmissions follow the same pattern as the 4,000-mc transmission; thus, the analysis of the test results can be made in the normal manner. The parameter *k* being the ratio of the effective earth curvature to the true earth curvature is generally accepted as a means of expressing the amount of bending of a ray caused by atmospheric refraction. At 4,000 mc during daytime hours and uniform weather conditions, *k* seldom deviates appreciably from 4/3. However, if possible, the value of *k* should be determined at the time of each test to permit the proper analysis of the test results. Even though the tests discussed in this paper were not conclusive because of the lack of obstructions toward the center of the path, where a change in bending is the greatest, the indications are that there is a very slight decrease in *k* when the frequency is increased. However, as the change is so small, the same value of *k* can for all practical purposes be used with these frequencies when determining an-

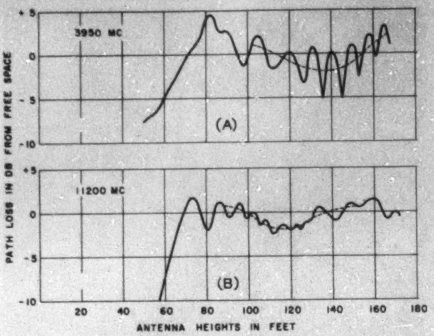


Fig. 4. Tests between Church Butte and Evanston

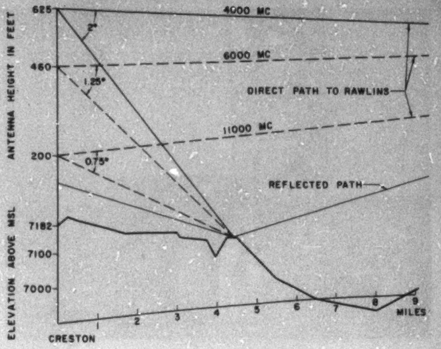


Fig. 5. Profile near Creston

tenna heights, except when severe multiple reflections are present in a path.

The general rule for microwave paths is that when no reflections are present and the clearance between the direct path and the highest obstruction is 0.6 or more of the first Fresnel zone radius, a signal close to the computed free space signal is received. The Fresnel zone radius is a function of frequency. When the frequency is increased, the radius becomes smaller, and thus less path clearance is required. An antenna fixed at the height needed for 4,000-mc transmission will therefore be satisfactory at 6,000 and 11,000 mc as the clearance will be greater than required.

When the frequency is increased, the gain of the antenna is also increased. This results in a narrower beam width, which gives greater discrimination against undesirable nearby reflections. This is evidenced in the test results by a decrease in amplitude of peaks and dips caused by the odd and even Fresnel zone effects. Fig. 4(A) shows the result of a 3,950-mc test between Church Butte and Evanston. Fig. 4(B) is an identical test made simultaneously at 11,200 mc. The long, sweeping valley which occurs at about 140 feet in Fig. 4 (A) and 117 feet in (B) is caused by an even Fresnel zone reflection from a surface 12.5 miles from Evanston. Thus the angle between the direct and reflected paths is small and there is no antenna discrimination at either frequency. The maximum reduction in signal is about the same: 2 db (decibels) below free space. The closely spaced peaks and dips are caused by reflections from surfaces near both ends of the path. At 4,000 mc, Fig. 4(A), the difference between the maxima and minima reaches as high as 6 db, whereas at 11,000 mc, (B), it is only about 2 db on account of the greater antenna discrimination.

In paths with reflections the advantage of greater antenna discrimination at the higher frequencies can best be illustrated by the following example. In the Creston to Rawlins path the location of one major reflecting surface is about 4.5 miles from Creston and its ground elevation is 100 feet lower than at this site. With vertical polarization the vertical beam width of a Western Electric Company horn reflector antenna is 2.0 degrees at 3,740 mc, 1.25 degrees at 6,325 mc, and 0.75 degrees at 10,960 mc.

To obtain half-power discrimination against reflections from the aforementioned surface, a 625-foot tower would be required at Creston to obtain an angle of 2.0 degrees between the direct and the reflected rays with a frequency of about 4,000 mc. At 6,000 mc a 460-foot tower would suffice to give an angle of 1.25 degrees, and at 11,000 mc a 200-foot tower will give 0.75 degree. This is shown in Fig. 5.

The fact that the antenna discrimination increases with frequency tends to give an impression that any path suitable for 4,000-mc transmission can also be used for the higher frequencies from the point of view of reflections. For several reasons this is an erroneous impression; as a matter of fact, any path showing evidence of severe reflections must for many reasons be carefully studied before its suitability can be determined for the frequencies in question. For instance, often the 4,000-mc antennas are mounted low behind a ridge so that the latter will block undesirable reflections from nearby surfaces beyond the ridge, and also so that the clearance between the direct path and the ridge remains within the limits of the first Fresnel zone for any value of k between 2/3 and infinity. This is done in order to prevent undesirable coinciding even-zone reflections from reaching the receiving antenna. At the Creston site, for instance, the towers were erected in a gully as shown in Fig. 6. The antennas facing Rawlins were located at such a height, 31 feet, that the clearance between the direct path and the ridge during normal transmission conditions is about equal to the radius of the first Fresnel zone. This is illustrated with a dashed line in Fig. 7(A) which shows a height-loss curve taken at 3,950 mc. The tests were made when $k=4/3$. The total apparent change in earth curvature is only ±3.25 feet at the ridge, when the value of k drops to 2/3 or rises to infinity. Thus, it is evident that the received signal will not drop below free space. Fig. 7(B) shows the results of an identical test made at 6,310 mc. Apparently the same antenna height can be used at Creston.

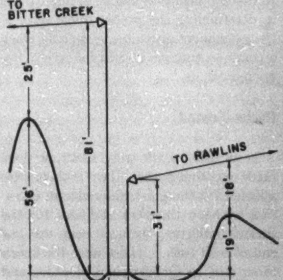


Fig. 6. Antenna arrangement at Creston

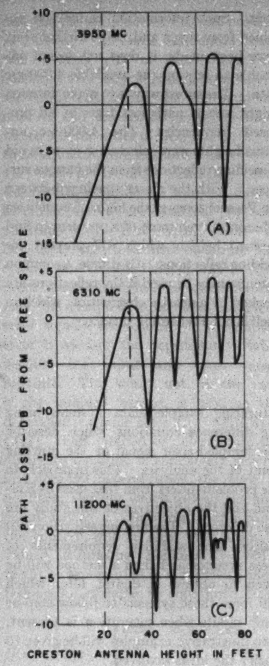


Fig. 7. Tests between Creston and Rawlins

The results of the 11,200-mc tests are shown in Fig. 7(C). If a common antenna is used, the received signal will even under normal transmission conditions be influenced by the second Fresnel zone reflection. During an increase in k, the loss will be even greater, especially at times when the ground is covered by snow or ice. If the antenna is lowered to 28 feet above ground to accommodate the 11,000-mc band, it may not affect the 4,000- and 6,000-mc transmissions severely, but the clearance above the ridge is reduced to such an extent that any vehicle passing on the road which runs along the ridge will cause a loss in signal. Furthermore, the physical dimensions of the horn reflector antenna prevent such a low antenna height. A solution to the problem is to locate separate 11,000-mc antennas at an elevation of 200 feet or more. At this height the angle between the direct and the reflected paths will exceed the antenna discrimination angle sufficiently to prevent severe signal cancellations.

Possible variations in reflection coefficient with frequency could not be determined from these tests in the absence of accurate information with respect to the performance characteristics of the test antennas at the higher frequencies. Tests using 6,310 mc were conducted between Buckhorn Mountain and Greeley using both horizontal and vertical polarization. There was no noticeable difference in the results. The peaks and dips in signal strength caused by the ground reflections were of the same magnitude, indicating that the phase angle and the reflection coefficient from the bare fields were the same for both types of polarization. This, however, may not be true in other terrain and under other circumstances. The effect of precipitation at the higher frequencies could not be studied as the weather conditions remained favorable throughout the tests.

Conclusions

The purpose of these tests was to collect certain information which will permit determining from earlier 4,000-mc tests the feasibility of using a common antenna for 4,000-, 6,000-, and 11,000-mc transmissions over existing paths. The conclusions drawn from the tests are as follows.

In paths which show no evidence of undesirable reflections there appear to be no other restrictions for using a common antenna than possibly the path length. Some existing 4,000-mc paths being up to 50 miles long may prohibit the use of 11,000 mc as a result of additional attenuation caused by precipitation.

Paths with only one reflecting surface will require a study of the 4,000-mc tests to determine if they are suitable for 6,000 and 11,000 mc. If the reflections are severe, the 4,000-mc antennas are preferably placed at such heights that the clearance between the direct radio path and the surface reamins greater than 0.5 but never more than 1.25 of the first Fresnel zone radius, as k varies from 2/3 to infinity. Because of the smaller Fresnel zone radii at 6,000 and 11,000 mc, the first zones will not occur at the same antenna heights and therefore the antennas must be placed at lower heights to avoid signal cancellation due to second-zone reflections. Furthermore, if the reflecting surface is located toward the center of the path, where the effect of atmospheric refraction is the greatest, the radio paths for the higher frequencies will not remain within these limits as k varies, even if separate antennas are used at different heights.

In paths with a reflecting surface close to one end the temptation is to orient vertically for the maximum signal. Regardless of frequency this will result in a very poor signal when a change in atmospheric refraction places a beam in an even zone above the reflecting surface, causing maximum cancellation in signal. When a common antenna is used, the higher frequencies will be even more vulnerable to misorientation, since the beam width becomes narrower. This can be illustrated by the following examples. Assume that the reflection coefficient is unity; the angle between the direct and reflected paths at the antenna nearest the reflecting surface is 2 degrees; and the antennas are oriented for maximum received signal. In the first example the heights of the antennas are such that the direct beam is under the influence of an odd Fresnel zone reflection. At 4,000 mc the received signal will have a broad peak with a maximum when the antenna is pointed one degree below the direct path. At 11,000 mc the signal will suffer a considerable loss, since the half-power beam width is only 0.75 degree and the antenna is oriented one degree away from either the direct or the reflected paths. In the next example it is assumed that the antennas are set at such heights that the direct beam is under the influence of an even Fresnel zone reflection. If they are oriented for the upper maximum, the antenna nearest the reflecting surface will be pointing above the direct path. Again the 11,000-mc signal will suffer some loss but it will not be as severe as in the previous example, since the angle to the direct path will be considerably smaller than one degree and at this frequency the antenna will discriminate against the reflected signals. No doubt, discrimination against undesirable reflected signals can be obtained by vertical orientation of an antenna and this seems to point toward a solution to the problem, but exaggerated misorientation of an antenna may severely reduce its ability to discriminate against signals polarized 90 degrees with respect to each other.

A common antenna may be used for paths where the existing antenna heights are such that the reflecting surface is affectively blocked by nonreflecting obstructions, or where the angle between the direct and reflected paths is so large that the antenna discriminates against the reflections. If these conditions are not fulfilled it may require separate antennas for satisfactory transmission. Paths over smooth earth such as water and salt flats are special cases which have so far, been tested with 4,000 mc only. Probably the reflections will affect the 6,000- and 11,000-mc transmissions in a similar manner, so the proper antenna heights can be calculated from the existing 4,000-mc data. Nevertheless it seems advisable to conduct some tests to obtain further knowledge about matters pertaining to this type of terrain and the effect of, for instance, salt storms on the path attenuation. Often the antenna at one end of a path is located several hundred feet higher than the one at the low end, which is only a few feet above ground, to keep the direct radio path within the limits of the first Fresnel zone. In such cases it may become physically impossible to mount a 6,000- or 11,000-mc antenna low enough.

Paths with multiple reflecting surfaces present a more complicated problem. The loss to the direct signal contributed by even Fresnel zone reflections from each surface separately may not be severe, but if the antenna heights are not properly selected the reflections from two or more surfaces will no doubt coincide during some bending condition. The combined effect can be severe enough to make the circuit unreliable. To predict without testing how the 6,000- and 11,000-mc transmissions will behave at present antenna heights is very difficult, if not impossible. It is necessary to have a complete and accurate knowledge of the paths with respect to the location and elevation of all reflecting surfaces, as well as obstructions. Such information cannot be obtained from maps and, until further tests have been made, it does not seeem advisable to rely on the available 4,000-mc data. The surest way is to make antenna height versus path loss tests at all proposed frequencies. The 4,000-mc antenna heights were selected so as to avoid coinciding reflections from the various surfaces. With the closer spacings between the Fresnel zones at the higher frequencies it becomes even more of a problem to find antenna heights which will prevent coinciding reflections. Of course, a common antenna may be used if the reflections are blocked by a nearby obstruction, which in itself is not a reflecting surface.

The Stability and Compensation of Saturating Sampled-Data Systems

FRANCIS J. MULLIN
ASSOCIATE MEMBER AIEE

Synopsis: Sampled-data feedback systems which contain a saturating element are investigated in this paper. A condition which must be satisfied if these systems are to be stable is developed; it is shown that this is a necessary condition for stability and an intuitive argument is presented to indicate that this condition is probably sufficient. Finally, a method to obtain a digital filter for compensating and/or stabilizing these systems is presented.

THE SYSTEM to be investigated in this paper is shown in Fig. 1. In particular it is wished to determine the effect of the saturating element on the response of the system when a step function is applied to the input.

Saturating sampled-data systems have been investigated previously by Nease[1] and Kalman.[2] In his study Nease considered pulse-width modulators in which the duration of the pulse applied to the system at the *n*th sampling instant was related to the error signal at the *n*th sampling instant. Since the pulse width can be no larger (in time) than the sampling period, a saturation effect between the pulse width and the sampled-error signal was necessarily introduced in his analysis. Kalman, on the other hand, discussed saturating systems in which the pulse width of sampled-error signal was small compared to the time constants of the system in which case the pulse could be accurately replaced by an impulse. This same assumption is made when the *z*-transform[3,4] or modified *z*-transform[5] are used to analyze linear sampled-data systems. Kalman was not primarily concerned with the analysis of these systems but rather with the problems involved in controlling such systems in some optimum fashion. Consequently, he talked about these systems in rather general terms and presented some concepts which can be used in finding this controlling function.

In this paper the study will be these saturating sampled-data systems using the difference equations which describe the sampled-error signal as the starting point of the analysis. This approach to the problem differs from that used by Kalman and the objectives of this study are also different. The first investigation will concern the stability requirements of these systems and then a method will be given to obtain a digital filter which will force these systems to follow step or ramp inputs when saturation is present. Two illustrative examples will be given to clarify the ideas involved.

Requirements for Stability

For this study the most convenient manner to describe the system is in terms of the sampled-error signal, $E(n)$, and its various derivatives. For notational convenience we shall use the following definitions (see Fig. 1):

$$\begin{aligned} E(n)&=e(t)\big|_{t=nT} & C(n)&=c(t)\big|_{t=nT} \\ \dot{E}(n)&=\frac{d[e(t)]}{dt}\bigg|_{t=nT} & \dot{C}(n)&=\frac{d[c(t)]}{dt}\bigg|_{t=nT} \qquad (1) \\ \ddot{E}(n)&=\frac{d^2[e(t)]}{dt^2}\bigg|_{t=nT} & \ddot{C}(n)&=\frac{d^2[c(t)]}{dt^2}\bigg|_{t=nT} \\ &\cdots & &\cdots \end{aligned}$$

$$R(n)=r(t)\big|_{t=nT}$$

T is the sampling period of the system and *n*, the independent variable of the difference equations, numbers the sampling period.

To form the difference equations which

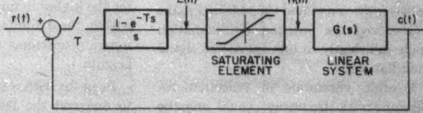


Fig. 1. A saturating sampled-data feedback system

Paper 59-52, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 30, 1958; made available for printing November 25, 1958.

FRANCIS J. MULLIN is with the California Institute of Technology, Pasadena, Calif.

This research was supported by the U. S. Air Force through the Air Force Office of Scientific Research, Air Research and Development Command under Contract no. *AF 18(600)-1521* at the University of California, Berkeley. The author would like to extend his personal thanks to Professors E. I. Jury and A. M. Hopkin of the University for their help and encouragement during the course of this research.

Vol.
Pt.
195

describe the sampled-error signal, proceed as follows. The linear differential equation relating $c(t)$ to $H(n)$, which is constant over any sampling interval, is solved for $c(t)$ during the nth sampling interval and differentiated $(p-1)$ times where p is the order of $G(s)$ in Fig. 1. If these equations are evaluated for $t=(n+1)T$, then p first-order difference equations which describe the sampled output in terms of $H(n)$ are obtained. By substituting $C(n)=R(n)-E(n)$, $\dot{C}(n)=\dot{R}(n)-\dot{E}(n)$, etc., into these equations and noting that $H(n)=f[E(n)]$, the p first-order difference equations which involve only the sampled input and sampled-error signals are formed. The sampled error is used as the dependent variable because the input and error terms separate. This would not be the case if the sampled output is used as the dependent variable since $H(n)=f[R(n)-C(n)]$. This method of determining the difference equation is completely general and does not depend upon the type of nonlinear element which follows the hold circuit in Fig. 1. However, in this paper the concern will be only with a single-valued saturating nonlinearity which might be represented in a piecewise linear fashion, for example, by equation 2.

$$H(n)=f[E(n)]=\begin{cases} KE_0; & E(n)>KE_0 \\ KE(n); & -KE_0<E(n)<KE_0 \\ -KE_0; & E(n)<-KE_0 \end{cases} \qquad (2)$$

K is the gain of the linear region of the saturating element. Assume that the linear part of the system, i.e., $G(s)$ in Fig. 1, has at least one but not more than $(p-1)$ integrators. This means that the step-function response of the linear part of the system of Fig. 1 may contain terms involving $t, t^2, t^3 \ldots t^K$, but $K<(p-1)$. Then, under these conditions, for any finite sampling period, it can be shown[6] that the magnitude of the sampled-error signal is bounded. This means that the sampled-error signal must have one of two properties: 1. it can approach a constant value, i.e., an equilibrium point in the phase space, and 2. it can approach a periodic solution in the phase space, i.e., a limit cycle.

Due to the assumption that $G(s)$ possesses at least one integrator, it is known that if the sampled-error signal approaches a constant value, this constant value is zero. Also known is the fact that if any oscillations are to occur then these oscillations must take place about the origin. Again, the reason for this is that there is at least one integrator in $G(s)$ and feedback from output to input.

The system is defined stable if, to a bounded input, there corresponds a bounded output, and the system is not capable of maintaining self-sustained oscillations. It has been indicated that the error signal (or output) is bounded and we would now like to develop a condition which must be satisfied if no limit cycle is to exist in the saturated system. For the systems investigated the difference equations which describe the sampled-error signal are:

$$E(n+1)=[f_{11}(E(n),T)]E(n)+f_{12}(T)\dot{E}(n)+f_{13}(T)\ddot{E}(n)+\ldots \qquad (3A)$$

$$\dot{E}(n+1)=[f_{21}(E(n),T)]E(n)+f_{22}(T)\dot{E}(n)+f_{23}(T)\ddot{E}(n)+\ldots \qquad (3B)$$

$$\ddot{E}(n+1)=[f_{31}(E(n),T)]E(n)+f_{32}(T)\dot{E}(n)+f_{33}(T)\ddot{E}(n)+\ldots \qquad (3C)$$

In these equations the nonlinear terms appear only as coefficients of $E(n)$ since there is an amplitude-dependent nonlinearity; the other coefficients are functions of only T, the sampling period. Also, the equations are not symmetrical, i.e., $f_{ij}(T)\neq f_{ji}(T)$. If the system is to be stable, then there exists a steady-state value of the error signal (which for systems with integrators is zero) called $E_{ss}(n)$, which can be obtained from equation 3(A) with $E(n+1)=E(n)$ and all derivative terms set equal to zero. Imagine that the system has reached this steady-state value. Then, if $E_{ss}(n)$ is to be an equilibrium point, any perturbation about $E_{ss}(n)$ must result in decaying oscillations; that is, the operating point of the system must eventually return to $E_{ss}(n)$ if this operating point is changed slightly from its steady-state value. Now, if the perturbed values of $E(n)$, $\dot{E}(n)$, etc., are denoted by $e(n)$, $\dot{e}(n)$, etc., then to a first approximation, the difference equations for the perturbed error are:[1]

$$e(n+1)=\overline{f_{11}}(T,E_{ss})e(n)+f_{12}(T)\dot{e}(n)+f_{13}(T)\ddot{e}(n)+\ldots \qquad (4A)$$

$$\dot{e}(n+1)=\overline{f_{21}}(T,E_{ss})e(n)+f_{22}(T)\dot{e}(n)+f_{23}(T)\ddot{e}(n)+\ldots \qquad (4B)$$

$$\ddot{e}(n+1)=\overline{f_{31}}(T,E_{ss})e(n)+f_{32}(T)\dot{e}(n)+f_{33}(T)\ddot{e}(n)+\ldots \qquad (4C)$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

where

$$\overline{f_{1j}}=\frac{\partial}{\partial E(n)}[f_{1j}(E(n),T)E(n)]\Big|_{E(n)=E_{ss}} \qquad j=1,2\ldots p \qquad (5)$$

Equations 4(A–C) are p first-order difference equations which can be solved for the perturbed error, $e(n)$, in a straightforward manner.[7] If the solution for $e(n)$ shows that the perturbation decreases with increasing n, then $E_{ss}(n)$ is a stable equilibrium point. On the other hand, if the solution shows that the perturbation does not decay with increasing n, then $E_{ss}(n)$ is an unstable point and a limit cycle will exist. This is certainly a necessary condition for the existence of a limit cycle.

We would now like to indicate the plausibility of this being a sufficient condition although this cannot be proved in a rigorous fashion. In the equations for the perturbed error signal 4(A–C) the quantities $\overline{f_{1j}}(T, E_{ss})$ give the slope of the saturation curve at E_{ss}. This means that the equations for the perturbed error signal are simply the nonlinear difference equations which describe the system, linearized in the vicinity of E_{ss}. Also, the equations for the perturbed error signal are the same as those of the completely linearized system where this system is one in which the saturation effect is ignored and the saturating element is assumed to have a gain of K over an indefinite range of input magnitudes. If the actual saturation curve has its maximum slope or highest gain in the vicinity of the origin, then the equations for the perturbed error signal are linearized in the vicinity of the highest gain. Now, if the completely linearized system is stable for this maximum gain, the equations for the perturbed error signal indicate decaying oscillations since these equations are the same as those of the completely linearized system. From a stability point of view the completely linearized system is "worse" than the saturated system because there are no bounds placed on the magnitude of the error signal. Consequently, it is felt that if the system is stable when the error signal is not bounded, it should be stable when this signal is bounded. And because the equations for the perturbed error signal are the same as the equations for the completely linearized system, it is believed that the previously stated condition is also a sufficient condition for stability.

In other words, the original nonlinear system has been replaced with a linear system having a gain equal to the maximum gain of the saturating system. Then if the linear system, which places no bound on the magnitude of the various signals in this system, is stable, it seems as though the saturated system, which restricts the magnitude of the various signals, should also be stable since the equations for the completely linearized system are the same as those for the perturbed error of the saturating system. The definite exceptions to this statement are those systems which are conditionally stable, in which case a root-locus plot of the poles of the completely linearized sys-

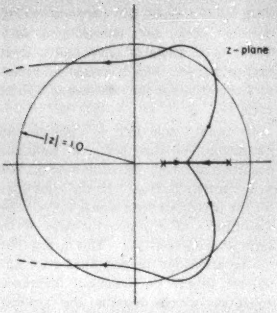


Fig. 2. A root locus plot of two poles of a conditionally stable sampled-data system. Arrows indicate the direction of increasing gain

tem is outside of the unit circle in the z-plane for certain values of gain but inside the unit circle for larger values of gain. Such a root-locus plot might be as shown in Fig. 2 and it seems likely that such a condition would occur only for very small values of T.

The summary of the stability criterion for these systems might be stated in the following manner. A necessary condition for saturating sampled-data systems which contain at least one integrator to be stable (no limit cycle exists) is that the equations for the perturbed error signal show that the perturbation decreases with increasing n. It seems plausible that this is a sufficient condition but this could not be proved rigorously.

Compensation Procedure

The system to be compensated is that of Fig. 1. It is wished to find a digital filter or discrete compensator which will operate on the sampled-error signal in such a fashion as to force the error signal to zero for inputs which are step functions of restricted magnitude or ramps with restricted slopes. In block diagram terminology this means that we want to find the $D^*(z)$ of Fig. 3 which will force the error to zero, subject to the condition that $D^*(z)$ be physically realizable. By physically realizable[8] is meant that the roots of the denominator of $D^*(z)$ must lie inside the unit circle in the z-plane and that the highest power of z in the denominator of $D^*(z)$ must be equal to, or greater than, the highest power of z in the numerator.

For simplicity, assume that what is wanted is to find a discrete compensator which will force the system to follow step functions with a restricted range of magnitudes. Apply a step function equal to the saturation level of the amplifier to the linear part of the system and find the time required for its output to reach a value equal to the maximum step function the system is to follow. This time places a lower bound on the time or number of sampling periods required for the error to go to zero for the entire range of input magnitudes and determines the minimum number of sampling periods which must elapse before the output and input can be equal for all values of the input. Now since this minimum time was determined by applying the maximum step function possible to the linear part of the system, it is obvious that the output of the system cannot be made equal to the input for all time in this minimum time. The time required to do this must be greater than this minimum time since a negative signal must be applied to the linear part of the system to keep the output from overshooting its desired value. In other words, if the system is initially driven in one direction with a series of positive pulses, then the only means of bringing the output to some constant value is to apply a series of negative pulses to the system later. If these later pulses are properly chosen it is possible to force the output to be equal to the input for all time after the second pulse sequence is applied. Thus, the initial problem which must be solved is to find a series of pulses of duration T which will cause the output and input to be the same within a reasonable length of time. The duration of these pulses must be T because the error signal of the system can change only at the sampling instants, and the magnitudes of these pulses must be less than the saturation level of the amplifier. This second condition is the one which causes the difficulty and makes this problem different from that of finding a discrete compensator for a linear system.

The series of pulses required to force the output to equal the input can be found as follows. Consider the sequence of pulses which are applied to the input of the linear element to be composed of a series of delayed step functions as shown in Fig. 4. Then, using the delayed step functions as the input to the linear part of the system, the output of the linear element can be written:

$$c(t)=\sum_{i=0}^{K} H_i h(t-iT) \tag{6}$$

where $h(t)$ is the step-function response of the linear system. The number of step functions, $(K+1)$, is determined from a consideration of the range-input magnitudes which the system should follow and it is possible to determine the range of input magnitudes for a given K. This will be pointed out in example II. This equation for $c(t)$ will contain terms which involve t and terms which do not involve t. The quantities which involve t are equated to zero for $t>KT$ and the sum of the constant terms is equated to the

Vol.
Pt. 1
1959

Fig. 3 (below). A block diagram showing the position of the discrete compensator, $D^*(z)$

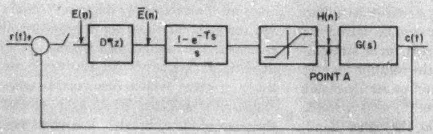


Fig. 4 (right). (A) The compensating step-function sequence at point A. (B) The equivalent pulse sequence at point A, $H(0)=H_0$, $H(1)=H_0+H_1$, $H(2)=H_0+H_1+H_2$, etc.

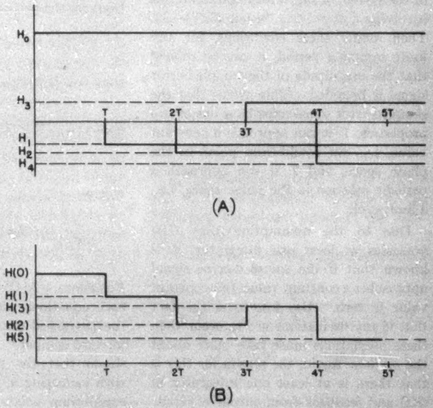
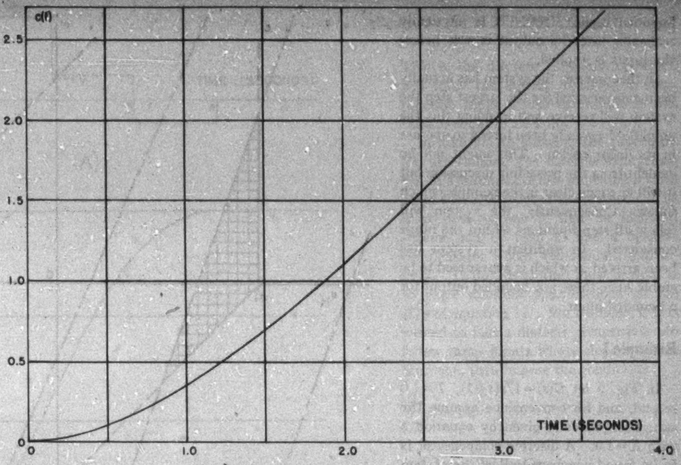


Fig. 5. The unit step-function response of $G(s)=1/s(s+1)$



value of the largest input step function the system will receive. This gives the relationship between the H_i's necessary to force the output to be equal to the input. There will always be more unknown H_i's than there are equations indicating that some of the H_i's can be arbitrarily specified. For convenience the H_i's will be arbitrarily specified from H_0 onward. For example, if three values of H_i can be specified, then the three values chosen will be assigned to H_0, H_1, and H_2. Then solve for the unknown H_i's in terms of those specified as arbitrary.

The word arbitrary has been used loosely here since the condition of restricted magnitude must be imposed on the amplitude of the impulses. To find what must be done to satisfy this condition, let V denote the saturated output of the amplifier. Then, the restricted magnitude condition can be stated:

$$|H_0| \leqslant V \tag{7A}$$

$$|H_0+H_1| \leqslant V \tag{7B}$$

$$\cdots\cdots\cdots\cdots$$

$$|H_0+H_1+\ldots+H_K| \leqslant V \tag{7K}$$

Since all of the unspecified H_1's have been expressed in terms of those we are at liberty to specify, all of these $(K+1)$ equations can be written in terms of only those H_1's which are arbitrary. These equations take care of the magnitude restrictions and then simply find the unspecified H_1's which satisfy these magnitude conditions. After obtaining them, compute the other H_1's and obtain the desired sequence of pulses. This can be done in a straightforward manner.

Once the pulse sequence has been determined, the discrete compensator which will produce this sequence of pulses can very easily be found. Knowing the pulse sequence, the sampled output of the system and the signal $\bar{E}(n)$ of Fig. 3 may be found as also the sampled error, $E(n)$, since the sampled input and output are known. $E(n)$ is the input to $D^*(z)$ and $\bar{E}(n)$ is the output of $D^*(z)$. Both are discrete signals which are zero after a finite time. This means we can take the z-transform of these two signals and obtain

$$D^*(z)=\frac{\bar{E}^*(z)}{E^*(z)} \tag{8}$$

$D^*(z)$ will always be a stable device since it has a bounded output, $\bar{E}^*(z)$, for a

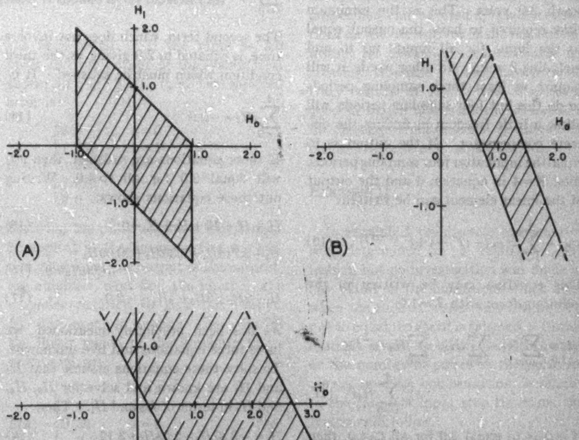


Fig. 6. (A) Magnitude restrictions imposed by equations 21 and 22. (B) Magnitude restriction imposed by equation 23. (C) Magnitude restriction imposed by equation 24

bounded input, $E^*(z)$. It is physically realizable since its output is zero before the input is applied.

In this process, the system has actually been compensated for the largest step the system will receive and in doing this the amplifier has really been forced to operate in its linear region. This might not be evident from the preceding discussion but it will be made clear in the examples which follow. Consequently, the system will follow all step functions within the range considered. In addition a system has been arrived at which is guaranteed to be stable since there is a bounded output for a bounded input.

Example I

In Fig. 3 let $G(s)=1/s(s+1)$, $T=1.0$ second, and for convenience assume the saturation curve is given by equation 2 with $K=1.0$. A discrete compensator is to be found which will allow the system to follow step inputs whose magnitudes may be as large as two units.

First find the response of $G(s)$ to the largest step function this linear network can receive. In this example it will be a unit step and for a unit step input, the output of $G(s)$ is

$$c(t)=t-1+e^{-t} \tag{9}$$

This response is shown in Fig. 5 and it is seen that it takes 2.95 seconds for $c(t)$ to reach 2.0 volts. This is the minimum time required to have the output equal to the input for all inputs up to and including 2 volts. In other words, it will require at least three sampling periods to do this but four sampling periods will allow a little freedom in finding the discrete compensator. If the output is to equal the input after four sampling periods, then $K=4$ in equation 6 and the output of the linear element can be written:

$$\underset{t\geq 4.0}{c(t)}=\sum_{i=0}^{4}H_i[(t-iT)-1+e^{-(t-iT)}] \tag{10}$$

This equation may be written in the following form with $T=1.0$

$$\underset{t\geq 4.0}{c(t)}=t\sum_{i=0}^{4}H_i-\sum_{i=0}^{4}iH_i-\sum_{i=0}^{4}H_i+e^{-t}\times\sum_{i=0}^{4}H_ie^i \tag{11}$$

If $c(t)$ is to equal 2.0 for all $t>4.0$, then the first and third terms of this equation require that

$$\sum_{i=0}^{4}H_i=0 \tag{12}$$

and the last term requires that

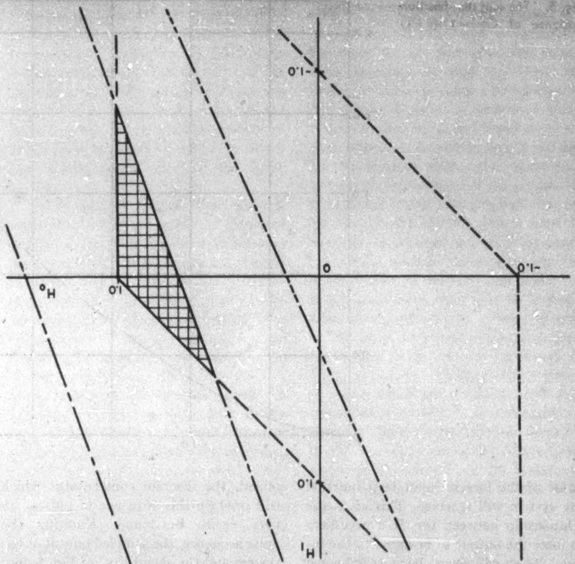


Fig. 7. The superposition of all magnitude restrictions imposed on H_0 and H_1. Shaded region shows permissible values

$$\sum_{i=0}^{4}H_ie^i=0 \tag{13}$$

The second term, which does not involve time, is equated to 2.0 giving us the third condition which must be satisfied. It is:

$$\sum_{i=0}^{4}iH_i=-2.0 \tag{14}$$

If these equations are satisfied, then $c(t)$ will equal 2.0 for all $t>4.0$. Writing out these equations shows:

$$H_0+H_1+H_2+H_3+H_4=0 \tag{15}$$

$$H_0+2.718H_1+7.39H_2+20.09H_3+54.60H_4=0 \tag{16}$$

$$H_1+2H_2+3H_3+4H_4=-2.0 \tag{17}$$

As has been previously mentioned, we have three equations and five unknowns. To solve these equations assume that H_0 and H_1 are known and solve for H_2, H_3, and H_4 in terms of H_0 and H_1. Then

$$H_2=-3.82H_0-2.03H_1+3.12 \tag{18}$$

$$H_3=3.67H_0+1.705H_1-4.25 \tag{19}$$

$$H_4=-0.86H_0-0.397H_1+1.14 \tag{20}$$

Now impose the magnitude restrictions. This is most easily done by drawing an H_0, H_1 plane and plotting the equations which must be satisfied if the magnitude of the pulses are to be limited. Refer now to Fig. 6. The first two magnitude conditions are:

$$|H_0|\leqslant 1.0 \tag{21}$$

$$|H_0+H_1|\leqslant 1.0 \tag{22}$$

Equation 21 eliminates that part of H_0, H_1 plane to the right of the line $H_0=1$ and the left of the line $H_0=-1.0$; equation 22 eliminates the section above the line $H_0+H_1=+1.0$ and below the line $H_0+H_1=-1.0$. This is shown in Fig. 6(A). The third magnitude condition is:

$$|H_0+H_1+H_2|\leqslant 1.0$$

Using equation 18 this becomes:

$$|-2.82H_0-1.03H_1+3.12|\leqslant 1.0 \tag{23}$$

Equation 23 eliminates the section to the right of the line $(-2.82\ H_0-1{,}03\ H_1+3.12=1.0)$ and to the left of the line $(-2.82\ H_0-1.03\ H_1+3.12=-1.0)$ and this restriction is shown in Fig. 6(B). The fourth condition is:

$$|H_0+H_1+H_2+H_3|\leqslant 1.0$$

And using equations 18 and 19 this becomes:

$$|0.86H_0+0.397H_1-1.14|\leqslant 1.0 \tag{24}$$

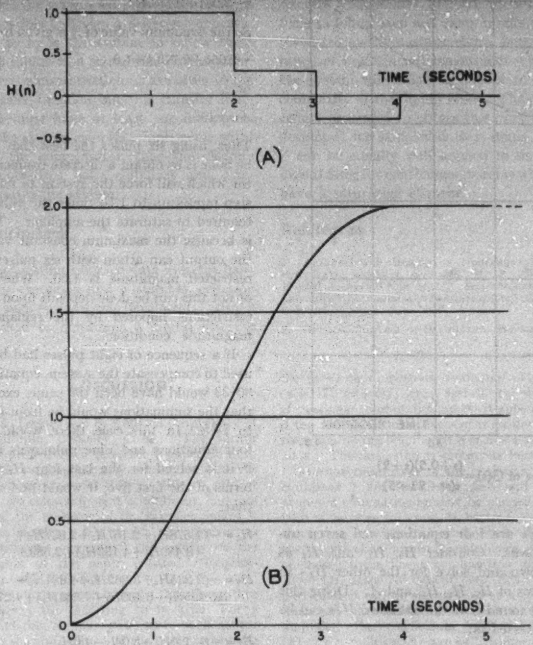


Fig. 8. (A) The pulse sequence to compensate $G(s)=1/s(s+1)$ with $T=1.0$ second. (B) The output of the system when this pulse sequence is applied to the input of $G(s)$

Equation 24 eliminates no additional region of H_0, H_1 plane as is seen in Fig. 6(C). The fifth magnitude condition is automatically satisfied since we have already imposed the condition $\Sigma H_i=0$ in determining the individual H_i's.

The result of all the magnitude restrictions is completely shown in Fig. 7. It is seen that any values of H_0 and H_1 bounded by the three lines ($H_0=1.0$), ($H_0+H_1=1.0$), and ($-2.82\ H_0-1.13\ H_1+3.12=-1.0$), will satisfy the magnitude restrictions and also force the output to equal the input in 4 seconds. To find the sequence of pulses let us arbitrarily select $H_0=1.0$ and $H_1=0$. Then $H_2=-0.70$; $H_3=-0.58$; and $H_4=0.28$. For these values of H_i, the desired pulse sequence is 1, 1, 0.30, −0.28, 0, 0, . . . This pulse sequence and the output of the linear system are shown in Fig. 8.

Since the output has been allowed to reach a value of 2.0, this is the response the system will have when the input is 2.0. Determining the discrete compensator which will give this response when the input is 2.0 units is straightforward now that the proper sequence of pulses has been found. If the input is 2.0 units and the output of the system is as shown in Fig. 8(B), then the sampled error is:

$$E(n)=\begin{cases}2.0 & n=0\\ 1.632 & n=1\\ 0.865 & n=2\\ 0.210 & n=3\\ 0.0 & n\geqslant 4\end{cases}\qquad(25)$$

Knowing the pulse sequence which must be applied to the linear part of the system, that is, at the output of the saturating amplifier, next find the input to the amplifier required to give this output. In this example assume the gain of the linear region of the amplifier to be unity and so:

$$E(n)=\begin{cases}1.0 & n=0\\ 1.0 & n=1\\ 0.30 & n=2\\ -0.28 & n=3\\ 0.0 & n\geqslant 4\end{cases}\qquad(26)$$

Using equation 8 it is finally found:

$$D^*(z)=\frac{1.0+1.0z^{-1}+0.30z^{-2}-0.28z^{-3}}{2.0+1.632z^{-1}+0.865^{-2}+0.21^{-3}}$$

or

$$D^*(z)=\frac{1.0z^3+1.0z^2+0.30z-0.28}{2.0z^3+1.632z^2+0.865z+0.21}\qquad(27)$$

This is the discrete compensator which will force the output to follow a step-function input whose magnitude is less than or equal to 2.0 units.

In addition to showing the procedure for determining the discrete compensator, this example has also indicated several other characteristics of the method. The most obvious one is that an identical scheme can be used to find a discrete compensator which will follow ramp inputs. To do this all that must be done is to place different specifications on the H_i's of equation 11. For example, if it is wished to find a discrete compensator to follow ramp inputs whose slope was less than one, then impose the conditions

$$\sum_{i=0}^{4}H_i=1.0$$

$$-\sum_{i=0}^{4}iH_i-\sum H_i=4.0$$

$$\sum H_ie^i=0$$

and find the H_i's as before, subject to the magnitude restrictions. With the H_i's determined, the discrete compensator is found exactly as it was for a step input.

Also, if the system has a time delay, this same method can be used both for step and ramp inputs. And, if it is desired, some of the H_i's which were left unspecified can be used to place restrictions on the derivatives of the response to insure, for example, a continuously increasing response. Thus, having arbitrary values of H_i at our disposal is good in that it allows additional freedom in meeting simultaneous specifications on the response and its derivatives.

Example II

In example I the discrete compensator was determined in a relatively easy fashion but no investigation was made in any detail on the limitations which arise when the time at which we wish the output to equal the input is selected. In this example it will be shown how this time, or the number of pulses in the pulse sequence used for compensation, is related to the range of input step functions the system will follow.

In Fig. 3 let $G(s)=(s+0.5)(s+2)/s(s^2+2s+s)$ and $T=1.0$ second. Also let the input-output characteristics of the amplifier be given again by equation 2 with $K=1.0$. If a unit step function is applied at the input of $G(s)$, then the output of $G(s)$ is:

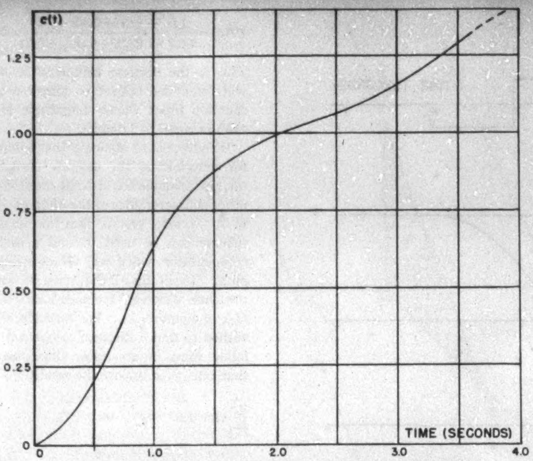


Fig. 9. The unit step-function response of $G(s)=\frac{(s+0.5)(s+2)}{s(s^2+2s+5)}$

$$c(t)=0.25+0.42-0.521e^{-t}\times \sin(2t+53.6° \text{ (degrees)}) \quad (28)$$

This response is shown in Fig. 9.

Use a sequence of six pulses to compensate the system and investigate exactly what can be done with those six pulses. With this input to $G(s)$ its output is:

$$\underset{t\geq 6.0}{c(t)}=\sum_{i=0}^{6} H_i[0.25(t-i)+0.42-0.521e^{-(t-i)}\sin[2(t-i)+53.6°]]$$

or

$$\underset{t\geq 6.0}{c(t)}=0.25t\sum_{i=0}^{6}H_i-0.25\sum_{i=0}^{6}iH_i+0.42\times\sum_{i=0}^{6}H_i-0.521e^{-t}\sum_{i=0}^{6}e^iH_i\times[\cos 2i\sin(2t+53.6°)-\sin 2i\cos(2t+53.6°)] \quad (29)$$

Now suppose that $c(t)$ = a constant, A, for $t>6.0$ and try to determine the maximum value of A. Then from equation 29 it can be seen that for this condition to be satisfied

$$\sum_{i=0}^{6}H_i=0 \quad (30)$$

$$-0.25\sum_{i=0}^{6}iH_i=A \quad (31)$$

$$\sum_{i=0}^{6}H_ie^i\cos 2i=0 \quad (32)$$

$$\sum_{i=0}^{6}H_ie^i\sin 2i=0 \quad (33)$$

Here are four equations and seven unknowns. Consider H_0, H_1, and H_2 as known and solve for the other H_i's in terms of H_0, H_1, H_2, and A. Doing this it is found that the unknown H_i's can be expressed as:

$$H_3=-1.693H_2-2.345H_1-3.051H_0+2.773A \quad (34)$$

$$H_4=-1.847H_3-2.622H_2-3.412H_1-4.216H_0+3.20A \quad (35)$$

$$H_5=2.0H_4-3.0H_3-4.0H_2-5.0H_1-6.0H_0+4.0A \quad (36)$$

$$H_6=-H_5-H_4-H_3-H_2-H_1-H_0 \quad (37)$$

In these equations specify H_0, H_1, and H_2. Then take $H_0=1.0$ to try to obtain a good initial response but let H_1 and H_2 remain arbitrary and assume that whatever these values are, they satisfy the magnitude conditions imposed by the saturation curve. This means that:

$$|1+H_1|\leqslant 1.0 \quad (38)$$

$$|1+H_1+H_2|\leqslant 1.0 \quad (39)$$

The fourth magnitude condition is:

$$|H_0+H_1+H_2+H_3|\leqslant 1.0$$

$$|-2.051-1.345H_1-0.693H_2+2.773A|\leqslant 1.0 \quad (40)$$

Now find values of H_1 and H_2 which will maximize A subject to equations 38 and 39. Since H_1 must be <0, the best choice of H_1 is $H_1=0$. Similarily if $H_1=0$, then $H_2<0$ and also is set $H_2=0$. This means:

$$|-2.051+2.773A|\leqslant 1.0 \quad (41)$$

So the maximum value of A is given by:

$$-2.051+2.773A=1.0 \quad (42)$$

or

$$A=1.10 \quad (43)$$

Thus, using six pulses the best that can be done is to obtain a discrete compensator which will force the system to follow step inputs up to 1.10 times the voltage required to saturate the amplifier. This is because the maximum constant value the output can attain with six pulses of restricted magnitude is 1.00. Whether or not this can be done depends upon the restrictions imposed by the remaining magnitude conditions.

If a sequence of eight pulses had been used to compensate the system, equations 30–33 would have been the same, except that the summations would be from $i=0$ to $i=8$. In this case there would be four equations and nine unknowns and if it is solved for the last four H_i's in terms of the first five, it would be found that:

$$H_5=-(1.578H_4+2.157H_3+2.837H_2+3.489H_1+4.133H_0)+2.583A \quad (44)$$

$$H_6=-(2.235H_5+3.262H_4+4.321H_3+5.395H_2+6.466H_1+7.534H_0)+4.275A \quad (45)$$

$$H_7=-(-2.0H_6-3.0H_5-4.0H_4-5.0H_3-6.0H_2-7.0H_1-8.0H_0)+4A \quad (46)$$

$$H_8=-(H_7+H_6+H_5+H_4+H_3+H_2+H_1+H_0) \quad (47)$$

Then, as was done with the six pulses, start finding the restrictions on the H_i's which the saturation effect causes. The maximum value of A will result if $H_0=1.0$, $H_1=H_2=H_3=H_4=0$ and then the sixth magnitude condition reduces to:

$$|H_0+H_5|=|1+H_5|\leqslant 1.0$$

or,

$$|-3.133+2.583A|\leqslant 1.0$$

This shows that the maximum value of A is given by:

$$A=\frac{4.133}{2.583}=1.60 \quad (48)$$

Thus, using an 8-pulse sequence, the best that can be done is obtain a discrete compensator which will force the system to follow step inputs up to 1.60 times the voltage at which the amplifier saturates. The remaining magnitude conditions will determine whether or not this is possible.

This example has shown how one can determine the range of input which can be followed for a given pulse sequence. This means that once the range of inputs which the system will receive is known, it is possible to design a discrete compensator to force the system to follow these step inputs in a straightforward manner. There is no restriction on the order of the transfer function but this transfer function must have at least one integrator. If the system exhibits a limit cycle without the compensator, then the discrete compensator both stabilizes and compensates the system.

Conclusions

Saturating sampled-data systems have been examined and a stability criterion for such systems was developed. Also, a method to determine a digital filter or discrete compensator to force these systems to follow step and ramp inputs was presented and it was also shown how the range of input signal magnitudes which the system will follow is related to the complexity of the digital filter. The procedure presented to obtain this filter was developed for sampled-data systems but it can be equally well applied to find a digital filter for continuous systems which have a saturating element.

References

1. Analysis and Design of Nonlinear Sampled-Data Control Systems, R. F. Nease. *WADC Technical Note 57-162*, Servomechanisms Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., June 1957.

2. Optimal Control of Saturating Systems by Intermittent Action, R. E. Kalman, *WESCON Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 4, 1957, pp. 130–35.

3. The Analysis of Sampled-Data Systems, J. R. Ragazzini, L. A. Zadeh. *AIEE Transactions*, vol. 71, pt. II, Nov. 1952, pp. 225–34.

4. Sampled-Data Control Systems, (book), E. I. Jury. John Wiley & Sons, Inc., New York, N. Y., 1958.

5. Synthesis and Critical Study of Sampled-Data Control Systems, Eliahu I. Jury, *AIEE Transactions*, vol. 75, pt. II, July 1956, pp. 141–51.

6. The Analysis and Compensation of Nonlinear Sampled-Data Feedback Systems, F. J. Mullin. *Ph.D. Thesis*, University of California, Berkeley, Calif., Sept. 1958.

7. A Note on the Operational Solution of Linear Difference Equations, E. I. Jury, F. J. Mullin. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 266, no. 3, Sept. 1958, pp. 189–205.

8. Discrete Compensation of Sampled-Data and Continuous Control Systems, Eliahu I. Jury, William Schroeder. *AIEE Transactions*, vol. 75, pt. II, 1956 (Jan. 1957 section), pp. 317–25.

◆

Discussion

C. A. Desoer and **S. Kodama** (University of California, Berkeley, Calif.): The author has presented an interesting paper on nonlinear sampled-data systems. He deserves commendation, for this is a difficult subject on two counts: first, it is nonlinear, and second, the presence of the sampler makes it time varying. We would like, however, to offer some comments* particularly on the section of the paper entitled, "Requirements for Stability."

From the paper it seems that the author believes the following to be true: For a saturating sampled-data servo such as the one shown in Fig. 1 of the paper, when the transfer function $G(s)$ is of the order p and contains at least one but not more than $(p-1)$ integrators then, for any finite sampling period, the magnitude of the sampled-error signal is bounded.

First, this theorem as it stands is misleading and may suggest incorrect conclusions, and second, a study of the author's proof (reference 6 of the paper) discloses some gaps in the reasoning.

That the theorem is misleading is best shown by a counter example:[1] Let the saturating amplifier have unity gain over its linear region and let $G(s)=1/s^2(s+1)$, then it can be shown that, if the input $r(t)$ remains equal to zero, any output phase-space trajectory which starts from the subspace defined by $c(nT)=E_0$, $\dot{c}(nT)\geq 5.5\ E_0$, $E_0\geq \ddot{c}(nT)>0$, becomes unbounded regardless of the sampling period. More precisely, it can be shown that the output behaves in an oscillatory manner; its amplitude and half-cycle period increase with time at least logarithmically.

Since the author's theorem leads to erroneous conclusions it should be useful to indicate where the proof's logic (reference 6 of the paper) is faulty. The author's method of proof consists of the following steps: 1. he *assumes* that all the derivatives of $e(t)$ of order $p-1$ and higher are bounded, 2. he uses equations 3(A), 3(B), 3(C), etc., to express $E(n+1)$ as a linear combination of the $(p-1)$th derivatives evaluated at nT, $(n+1)T$, $(n+2)T$, etc., and 3. the form of the linear combination makes it appear as if the assumed boundedness of these derivatives implies the boundedness of $E(n+1)$ for all n. This last step is erroneous because 1. in some cases some of the coefficients of equations 3 are identically zero and the proposed elimination cannot be carried out (this is the case, for example, where $G(s)$ has more than one integrator), 2. some of the coefficients of equations 3 (and, hence, some of those of the linear combination giving $E(n+1)$) actually depend on $E(n)$. In particular it can be shown that as a consequence of the saturation characteristic of the amplifier, the coefficients $f_{21}(E(n),T)$, $f_{11}(E(n),T)$. . . all go to zero as $E(n)$ becomes arbitrarily large. Since these coefficients appear in the denominator of the linear combination giving $E(n+1)$, the third step of the author's reasoning is incorrect.

In view of these considerations it should be worthwhile knowing a correct statement of sufficient conditions for boundedness of the output together with a proof that satisfies the objections just mentioned. First the statement:

For a sampled-data servo with a saturating amplifier such as that shown in Fig. 1, with the input $r(t)$ equal to zero for all $t>0$ and an arbitrary set of initial conditions (by "arbitrary set of initial conditions" is meant any set of initial conditions that can be reached by applying to the servo any continuous input $r(t)$.), the output $c(t)$ will be bounded for all t (including sampling instants) if 1. $G(s)$ has left half-plane poles and only one simple pole at $s=0$ with a positive residue, and 2. the degree of the numerator polynomial is at least one less than that of the denominator polynomial.

Note that the statement does not make any particular assumption regarding the shape of the amplifier-gain characteristic except that it is single valued and that it saturates, i.e., as $E(n)$ becomes large the amplifier output $h(t)$ becomes equal to $\pm H_0$, the saturating level of the amplifier, more precisely when $|e(t)|\geq E_0$, $h(t)=e(t)/|e(t)|\,H_0$.

The proof follows. If $g(t)$ is the unit impulse response of $G(s)$, then

$$c(t)=\int_{-\infty}^{+\infty}h(\tau)g(t-\tau)d\tau$$

Let p be the degree of the denominator polynomial of $G(s)$. Differentiating this expression k times with respect to t,

$$c^k(t)=\int_{-\infty}^{+\infty}h(\tau)g^k(t-\tau)d\tau \quad k=1,2\ldots p-1$$

where $g^k(t)$ denotes the kth derivative of $g(t)$ taken in the distribution sense[2] so that it may contain impulses, doublets, etc., at $t=0$. As a result of the piecewise continuous character of $h(\tau)$, the aforementioned equality may not have a meaning where $k\geq 2$ for $t=nT$ because distribution theory does not define the integral of an impulse times a function that is discontinuous at the impulse. In view of the assumptions on $G(s)$, $g(t)$ is of the form

$$g(t)=A+\sum_k\left(\sum_l A_{kl}\frac{t^l}{l!}\right)e^{p_k t} \tag{49}$$

where $Re(p_k)<0$ for all k's, hence all its derivatives are absolutely integrable, in particular,

$$\int_{-\infty}^{+\infty}|g^k(t)|dt<M_k \quad (k=1,2\ldots p-1)$$

where M_k is an appropriate positive constant. Since for all t, $|h(t)|\leq H_0$ we have

$$|c^k(t)|\leq H_0\int_{-\infty}^{+\infty}g^k(t-\tau)d\tau$$

and

$$|c^k(t)|<H_0M_k \quad (k=1,2\ldots p-1)$$

the inequality being valid for all t if $k=1$ but may not be valid at sampling instants for $k\geq 2$. It should be stressed that the uniform boundedness of all derivatives of the output $c(t)$ (except possibly at sampling instants) follows from the fact that $G(s)$ contains a single integrator and that its other poles lie inside the left-half plane.

It remains to show that $c(t)$ itself is uniformly bounded. If it can be proved that the servo cannot remain in saturation for a time longer than T_1 (where T_1 is a constant independent of the initial conditions and dependent only on the characteristics of the servo) then, the already proved boundedness of $\dot{c}(t)$ implies that $c(t)$ is bounded. In particular

---

* The results reported here came from research supported by the U. S. Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract no. *AF18(600)-1521*.

$$|c(t)| < E_0 + H_0 M_1 \frac{T_1}{2}$$

This inequality follows from the following geometric argument: When the servo enters saturation, $c(t) = E_0$. (Here we assume positive saturation; in case of negative saturation $c(t) = -E_0$ but the remainder of the argument follows with a few trivial changes.) The existence of T_1 means that during the following T_1 seconds the servo has come out of saturation. Therefore the largest excursion of $c(t)$ would occur if $c(t)$ would increase at the maximum rate for $T_1/2$ seconds and decrease at the maximum rate for the remaining $T_1/2$ seconds.

It remains to establish the existence of time T_1. Suppose the servo enters saturation $(c(t) > E_0)$ in the interval $(m-1)T < t \leq mT$ and is still in saturation at $t = mT$, then $c(mT) < E_0 + H_0 M_1 T$ since the slope of the curve $c(t)$ is always smaller than $H_0 M_1$. Thus, as long as the servo remains in saturation the following situation exists: 1. the input to $G(s)$ is $-H_0$, and 2. the initial values just after the sampling instant mT are such that $E_0 < c(mT) < E_0 + H_0 M_1 T$, $|c^k(mT+)| < H_0 M_k (k = 1, 2. \ . \ . p-1)$. The solution of the differential equation indicates that as long as the servo is in saturation and $t > mT$

$$c(t) = \sum_{k=0}^{p-1} c^k(mT+)\psi_k(t-mT) - H_0 \int_{mT}^{t} g(\tau - mT) d\tau \quad (50)$$

The pole locations of $G(s)$ require that the transients $\psi_k(t-mT)$ are of the same form as the right hand side of equation 49.

Therefore the function

$$f(t) = \sum_{k=0}^{m-1} c^k(mT+)\psi_k(t-mT)$$

is uniformly bounded for all $t > mT$ by a constant M which is independent of the initial conditions since the latter are uniformly bounded. Now, equation 50 cannot hold for arbitrarily large t since the integral $\int_{mT}^{t} g(\tau - mT) d\tau$, for large enough t, increases linearly with t. As a consequence $c(t)$ given by equation 50 will eventually become smaller than E_0 and the servo will come out of saturation. The time the servo will remain in saturation will be bounded above by $T_1 = T_0 + T$ where T_0 is the smallest number x such that

$$\int_0^x g(\tau) d\tau = \frac{M - E_0}{H_0}$$

From this expression it is apparent that T_1 depends exclusively on the characteristics of the servo and is independent of the initial conditions.

In short, it has been shown that, except maybe at the sampling instants $t = nT$, the phase-space point $[c(t), \dot{c}(t), \ddot{c}(t) . \ . \ . c^{p-1}(t)]$ remains inside the parallelpiped defined by the following inequalities

$$|c(t)| < E_0 + H_0 M_1 \frac{T_1}{2}$$

$$|c^k(t)| < H_0 M_k \qquad (k = 1, 2 . . . p-1)$$

These inequalities hold whatever might be the initial conditions. The only thing that was proved was the boundedness of $c(t)$ and its $(p-1)$ first derivatives. This does not imply that the servo is asymptotically stable which would require proving that $c(t)$ and its $(p-1)$ first derivatives go to zero, as $t \rightarrow \infty$.

References

1. **C. A. Desoer, S. Kodama.** (available from the authors).

2. Principles and Techniques of Applied Mathematics (book), **B. Friedman.** John Wiley & Sons, Inc., New York, N. Y., 1956, chap. III.

Francis J. Mullin: The author would like to thank Professor Desoer and Mr. Kodama for their discussion and certainly agrees with their results; namely, that if the sampled-data system shown in Fig. 1 contains a saturating element and $G(s)$ is a transfer function with *one* integrator and all of its remaining poles in the left-half s-plane, then the continuous output of the system is bounded for all time. This is in contrast to the statement in the paper which permitted $G(s)$ to contain $(p-1)$ integrators and still have a bounded output. This statement in the paper of the restrictions on $G(s)$ is incorrect. $G(s)$ may contain only a single integrator if the output of the saturating system is to be bounded.

The proof for the boundedness of the output which the discussers have presented is both understandable and clever. The idea of showing that the system cannot remain indefinitely in saturation is the key to their argument and the fact that this is true, definitely shows the boundedness of the output.

The argument which was originally used to show the boundedness of the output was based on the difference equations which describe the sampled-error signal. The sampled-error signal was expressed as a combination of sampled derivatives of the error signal and since all derivatives are bounded, it was believed the sampled-error signal was bounded. This is true only if $G(s)$ is a stable transfer function with a single integrator. If this condition is not satisfied, then it is not possible to express the sampled-error signal in terms of the derivative quantities because some coefficients of the difference equations are identically zero. This fact was overlooked in reference 6. At any rate, the question on the boundedness of the output of the system in Fig. 1 has been cleared up by the discussion and the author appreciates the efforts of the discussers and this contribution.

Synchronous Methods of Operation for Private Telegraph Networks

B. S. DIAMOND
MEMBER AIEE

THERE ARE MANY private telegraph networks in the United States serving industrial, commercial, financial, and transportation organizations, and used for the transmission of intracompany message traffic, data transmission, and telemetering. In size there is a wide range, from small networks with a home office and a single branch, to large and highly complex ones with tens of thousands of miles of facilities, hundreds of operating points, and weekly message files running into six figures. The majority of private networks lease both equipment and facilities. However some own their operating equipment and lease facilities, while others own both equipment and the plant that provides the facilities.

Most trunk circuits are routed via voice-frequency channels with either 150 or 170 cps (cycles per second) allocated to each channel. There are, however, a number of radio circuits and a few physical lines of moderate length. Local drops are usually metallic loops.

The standard method of operation is start-stop at either 60, 75, or 100 wpm (words per minute). The exceptions include some leased multiplex channels, an occasional morse circuit left over from the past, telautographs or facsimile for certain types of local service, and a few wide-band installations for data transmission. The advantages of start-stop operation are obvious. Installations are easy to make, convenient methods of switching are available, maintenance and planning problems are relatively simple, and, when traffic loads are light, ordinary typists can take the place of professional telegraphists.

In areas where lines or facilities are cheap, plentiful, and efficient, start-stop is undoubtedly the preferred method of operation. On the other hand, where circuits are costly, and accuracy and

Paper 59-61, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted October 14, 1958; made available for printing November 25, 1958.

B. S. Diamond is a telegraph consultant, Madison, N. J.

efficiency are important factors in the operation of long-distance trunks, synchronous methods may be preferable.

The purpose of this paper is to show that private networks may be able to obtain significant advantages by the use of synchronous methods on expensive trunk circuits and on circuits subject to transmission difficulties.

Comparison of Transmission Features

Possible Disadvantages

In comparison with synchronous methods start-stop has disadvantages, which can be quite serious, on trunk circuits where accuracy, efficiency, and costs are factors of importance. Consider the following:

Tolerance for Distortion

Under the most favorable circumstances for either method of operation, a start-stop receiver has only one half the tolerance for distortion possessed by a synchronous receiver. Practically, the differential is nearer three to one because of the necessity of providing, in start-stop working, for a possible difference of 1% between rate of modulation and rate of receiver selection, and for unavoidable suboptimum adjustments of start-stop operating equipment.

Accuracy

In the case of a medium subject to characteristic distortion plus interference, the total amount of telegraph distortion, as discussed in the Appendix I, may be more than twice as much as would be apparent with characteristic distortion alone. Where there is more than one source of interference, there is a tendency for an isolated "pile-up" to occur. If a pile-up of, say, 20% signal amplitude coincided with a crossover, and with a discrepancy between rate of modulation and rate of receiver selection, an error is liable to occur, particularly if there are any significant suboptimum adjustments of the operating equipment.

When trunk circuits are lengthy, and comprise two or more sections in tandem, the error-rate may be quite high. In one network where circuits were normally complex because of the switching system, observations indicated an error-rate of one per 1,000 characters. In many error checks observed by the writer on complex chains, comprising both start-stop and synchronous links, the majority of errors occurred on the start-stop links.

Code Efficiency

Except where crypto operation is required, synchronous operation normally requires 5 units only per character as compared to the 7.42 cycle of start-stop.

Efficiency of Utilization

With synchronous operation as a reference, start-stop will normally have an efficiency factor of less than 50% even on voice-frequency channels specifically designed for start-stop operation.

Through other media, because there is a choice of but three modulation rates with start-stop, the efficiency of utilization may be quite low. If the characteristics of a transmission channel are such that a modulation rate of 45.45 bauds is not practicable, there is not even a circuit if dependent on normal start-stop methods.

On the other hand, there are no definite lower or upper limits for synchronous operation. Present-day modulation rates range from a low of probably 9 bauds to a high of 300 bauds. At least one modern synchronous system is potentially capable of 16 channels at 100 wpm per channel which, with 6-unit operation, would result in a modulation rate of 960 bauds.

Table I. Wpm for Various Modulation Rates

Modulation Rate, Bauds	Pulse Length, Milliseconds	Wpm 5 Unit	Wpm 6 Unit	Wpm 7 Unit	Wpm 7.42 Unit	Wpm 8 Unit
10	100	20	16.67			
20	50	40	33.3			
30	33.3	60	50			
40	25	80	66.7			
45.45*	22	90.9	75.83	65	61.3	56.8
50	20	100	83.3			
56.88†	17.57	113.76	94.8	81.2	76.67	68.5
60	16.67	120	100			
74.2‡	13.47	148.4	123.7	106	100	92.75
75	13.3	150	125			
100	10	200	166.7			

* Rate for nominal 60 or 65 wpm, start-stop.

† Rate for nominal 75 wpm, start-stop.

‡ Rate for nominal 100 wpm, start-stop.

The telegraph companies and the armed forces of the United States do not consider the operation of major overseas trunks to be either satisfactory of efficient with start-stop methods and generally these are operated synchronously.

Private telegraph networks do not have the direct economic and operational pressures that make the telegraph companies and the armed forces operate their circuits with optimum efficiency. Where facilities are plentiful and cheap, an abstract utilization efficiency factor means little; where reliable, the advantage of synchronous equipment in working through adverse conditions is not so important.

However, the facilities used by private networks cannot always be considered as being either cheap or plentiful. The cost of a 1,000-mile lease is not low and, while it may normally be possible to lease a 61-wpm start-stop facility between almost any two cities in Canada and the United States, there could be considerable difficulty in getting one suitable for 100 wpm, start-stop; see Tables I and II.

Neither are private networks always free of adverse transmission conditions. In making the survey for this paper, two comments on the difficulties of working start-stop through heavy interference from power mains were heard and it may

Table II. Utilization Factors

Type of Circuit	Wpm	Band Width, Cps	Utilization Factor
Standard telegraph voice-frequency carrier channels	65	120	0.54
	75		0.625
	100		0.83
	65	150	0.43
	75		0.5
	100		0.67
	65	170	0.38
	75		0.44
	100		0.59
Ocean cable carrier, reference circuit between Newfoundland and Nova Scotia with normal signals	65	50	1.3
Ocean cable carrier, reference circuit between Newfoundland and Nova Scotia with suppressed singles	83.3	50	1.67
Ocean cable physical, normal signals	300	113	2.65
Ocean cable physical, suppressed singles	300	60	5.0
Ocean cable physical, 3-element cable code	50	11.6	4.3

For comparing utilization efficiencies of transmission channels, the rate of speed may be divided by band width. Dividing wpm by the band width in cps provides a convenient set of reference figures to indicate relative efficiency as far as the movement of traffic is concerned. This figure may be called the utilization factor.

be assumed that these were not isolated. Transmission channels derived from radio beams are sometimes so seriously afflicted with atmospherics and multipath effects that communication with normal start-stop methods of operations is submarginal.

Possible Advantages

Some examples of situations where private networks might gain significant advantages from synchronous equipment follow:

Where Faster Trunk Circuits Are Required

If a transmission channel has a rating of 45.45 bauds, start-stop, and telegraph distortion is no greater than 10%, synchronous operation at a modulation rate of 74.2 bauds should be practicable and the following working speeds possible:

Units	Wpm
5	148
6	123
7	106
8	93

If the channel has more than 10% distortion but is still error-free when working start-stop, a modulation rate of 56.88 bauds should be possible and the following speeds obtained:

Units	Wpm
5	114
6	95
7	81
8	71

The above estimates are based on actual experience with multiplex circuits routed via transmission channels set up for 45.45 bauds, start-stop.

For transmission channels having a rating of 75 or 100 wpm, start-stop, the amount of gain possible from synchronous operation will depend on the efficiency of intersection couplings plus the amount of interference within the transmission channel. Such channels are normally space the same as those with a rating of 61 wpm and the higher speed obtained by circuit refinements. Hence, the gain from a changeover to synchronous operation is unlikely to be as great proportionately as on a channel with a rating of 61 wpm.

Where, Without Necessarily Suffering from Congestion, Channels Are Required for Traffic Routing Purposes

With this problem, advantage could be taken of the characteristic ease and reliability of channel splitting, character by character, of synchronous methods of operation. At the same time, a worthwhile increase in speed may be obtainable. Possible channel speeds and patterns are shown in Table III.

Where Circuits Are Subject to Interference from Power Mains

Consider the case of a circuit subject to interference from power mains to the extent that normal start-stop working was impossible. The working band for a 61-wpm start-stop circuit normally extends to a frequency between 50 and 70 cps. Hence, discriminatory networks will be of little value unless the amount of characteristic distortion present was small.

On the other hand, a 60-wpm 6-unit synchronous circuit will have a modulation rate of 38 bauds and only requires frequency components extending to 27 cps. With over an octave of clearance, a 6-db (decibel) improvement in signal to interference ratio could be anticipated with a capacitance of suitable value, 12 db with a simple L/C (inductance capacitance) network, and, as far as the mains frequency was concerned, probably up to 40 db with a more complex filter although, with over 20 db attenuation on the mains frequency, low-frequency transients, from variation on the mains load, might become prominent.

As an example, it was found, in laboratory experiments, that a synchronous circuit at 45.45 bauds was free of errors 60-cps interference superimposed on a signal having about 25% characteristic distortion. The amplitude of the interference was about 25% that of the signal; total distortion was about 60%. Given a similar situation in the field, a circuit with a total of 60% telegraph distortion would not have been considered safe for traffic. However, with a simple filter, the interference could be reduced by nearly 12 db to a residue of, say, 8% signal amplitude. The total distortion would then have been reduced to, possibly, 40% and, worked synchronously, the circuit would have been a solid one.

Where Circuits Are Subject to Interference from Atmospherics and Multipath Effects

On a radio link, multipath effects and atmospherics may make straight start-stop operation unsatisfactory. Where this occurs, heavy filtering with well-damped networks, plus synchronous operation, may convert an unsatisfactory circuit into a good one. In extreme cases, where it could be justified economically and suitable technical personnel was available, most of the errors could be eliminated by an automatic "RQ" system. Theoretically, provided that minimum signal amplitude was adequate, interference can be integrated and a circuit obtained at some speed which, of course, could be on the low side.

Table III

Modulation Rate, Bauds	Code, No. of Units	Two Channels, Wpm	Three Channels, Wpm	Four Channels, Wpm
56.88	5	57.0	37.9	28.4
	6	47.4	31.6	23.4
74.2	5	73.2	49.5	39.5
	6	61.8	41.2	31.2

Equipment Problems for Private Networks

Any advantages that private networks might obtain in theory from the use of synchronous equipment are not of much practical use unless suitable equipment is available. Hence, some design criteria should be developed.

Until the pressure of circumstances stimulated development work in synchronous techniques during the late 1940's and early 1950's, multiplex sets had certain drawbacks. Rotary distributors were driven from fork contacts and fork frequencies were none too stable. Synchronism was obtained by stepping the brush-arms with mechanism that was seldom completely reliable and which frequently either missed or took double jumps. A transitory bias might cause a rush of corrections that could result in an out-of-phase. A gap of a few seconds would also result in an out-of-phase. When, on domestic lines, large numbers of relatively inexpensive voice-frequency telegraph channels became available for start-stop working, the old multiplexes were discarded.

Later, the increase in volume of overseas message traffic and the demand for leased channels, concurrent with the general adoption of frequency-shift keying methods on radio and the repeaterization of long submarine cables, stimulated developments in synchronous techniques and time-division multiplex.

As a result, modern multiplex equipment is very different from that in use prior to World War II. Frequency control and synchronizing networks are now electronic with a great improvement in stability. Out-of-phases, because of speed variation and short gaps, no longer occur. The use of 5-wire printers for traffic reception has been generally discontinued and local loops, both sending and receiving, are mostly start-stop.

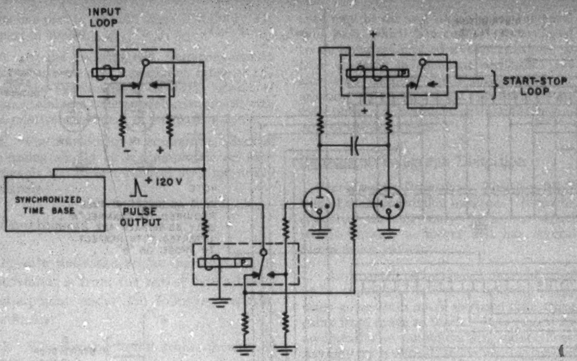

INPUT LOOP
-
+
SYNCHRONIZED TIME BASE
+120 V
PULSE OUTPUT
START-STOP LOOP
+
+

Fig. 1. Test regenerator

For coupling purposes, the modern multiplex channel is every bit as flexible as a start-stop one and possibly more convenient for certain methods of automatic switching.

However, the increasing emphasis on electronic methods has sometimes resulted in a tendency for complexity to increase and some of the more recent multiplexes, for one of which the writer has to accept responsibility, contain hundreds of electronic elements such as vacuum tubes, diodes, and transistors. With such complex equipment, a great deal of theoretical knowledge, experience, and study, together with ample patience and time, may be necessary to localize and clear an obscure dry joint, faulty connection, or defective circuit element.

Equipment of this type may be beyond the understanding of the most useful type of technician, the type indispensable for keeping a system running. That is, the good practical man with years of experience, who knows telegraph routine, has sufficient operating ability for, at least, the exchange of service messages, can maintain printers, and who has some technical training superimposed on a high school education.

To be of optimum value to private networks, synchronous equipment should be designed on the assumption that it will be in the hands of men with the knowledge, time, and experience required for the care of reasonably simple electronic circuitry but who might have difficulty with highly complex networks.

Methods of Operation

For single-channel synchronous operation, 5- or 6-unit operation would involve storage and the use of a translator for conversion from or to start-stop signals. If simplicity of equipment is considered to be an important factor, 7-unit operation, with start and stop pulses transmitted, might be considered preferable.

For single-channel 7-unit synchronous operation, rate of transmission would be controlled by a stable frequency with, at the receiving end, a synchronous regenerator followed by a teleprinter.

The normal 7.42 cycle is obviously impossible with a simple synchronous regenerator. However, the only reason for the prolonged stop pulse is to provide margin against, first, a possible speed differential between sending and receiving ends and secondly, line distortion. When the speed is stable and distortion cannot reach the printer, 7-unit operation should be satisfactory.

With 7-unit operation, a 60-wpm printer will work at 65 wpm; 75-wpm one at 81.2 wpm, and the 100-wpm type at 106 wpm. Unless carefully adjusted, the older types

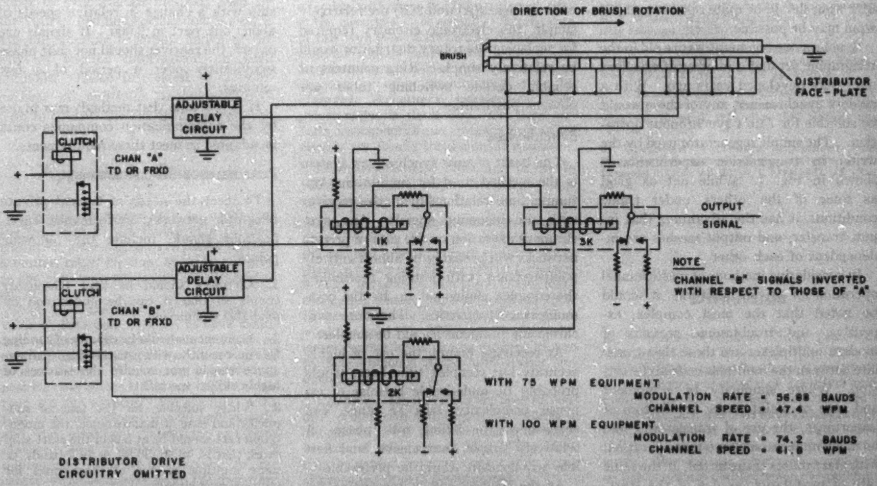

DIRECTION OF BRUSH ROTATION
BRUSH
DISTRIBUTOR FACE-PLATE
ADJUSTABLE DELAY CIRCUIT
CLUTCH
CHAN "A" TD OR FRXD
1K
3K
OUTPUT SIGNAL
NOTE
CHANNEL "B" SIGNALS INVERTED WITH RESPECT TO THOSE OF "A"
ADJUSTABLE DELAY CIRCUIT
CLUTCH
CHAN "B" TD OR FRXD
2K
WITH 75 WPM EQUIPMENT
MODULATION RATE = 56.88 BAUDS
CHANNEL SPEED = 47.4 WPM
WITH 100 WPM EQUIPMENT
MODULATION RATE = 74.2 BAUDS
CHANNEL SPEED = 61.8 WPM
DISTRIBUTOR DRIVE CIRCUITRY OMITTED

Fig. 2(A). Two-channel combining circuit

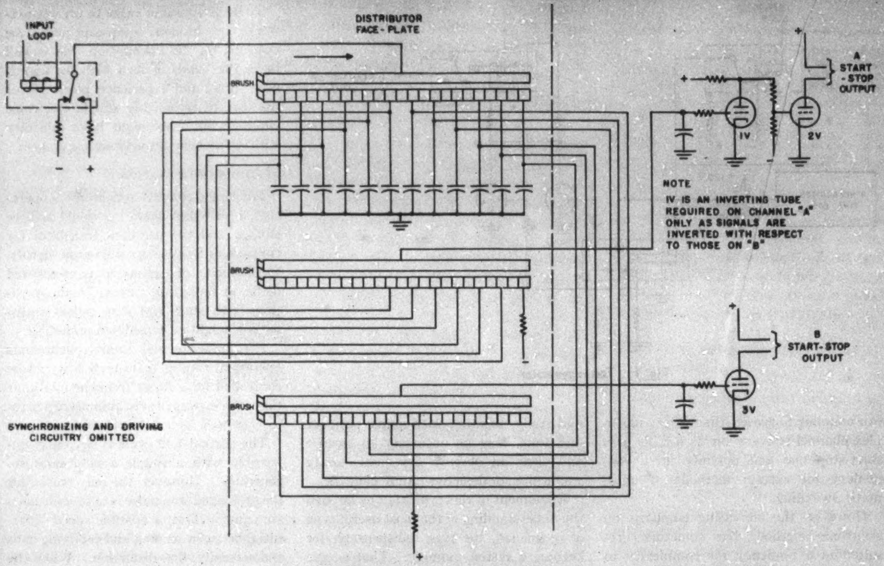

INPUT LOOP
DISTRIBUTOR FACE-PLATE
BRUSH
BRUSH
BRUSH
A START - STOP OUTPUT
1V
2V
NOTE
IV IS AN INVERTING TUBE REQUIRED ON CHANNEL "A" ONLY AS SIGNALS ARE INVERTED WITH RESPECT TO THOSE ON "B"
B START-STOP OUTPUT
3V
SYNCHRONIZING AND DRIVING CIRCUITRY OMITTED
+
-

Fig. 2(B). Two-channel receiving distributor

may not have quite enough margin for 7-unit operation at 100 wpm but the Model 28 has ample if signals are reasonably free of distortion. On a transmission line with a 60-wpm start-stop rating, 7-unit synchronous operation at 81.2 wpm should be quite reliable and 106 wpm may be possible.

Synchronous regenerators are old in the telegraphic art and a number of excellent ones were developed years ago. With a modern synchronizer, any of these would be suitable for 7-unit synchronous operation. The simple regenerator used by the writer in transmission experiments is shown in Fig. 1. While not as good as some of the others under traffic conditions, it has the advantage that input, transfer, and output meshes are independent of each other.

In considering methods of multichannel operation for private networks, it should be noted that the most complex, expensive, and troublesome sections of modern multiplexes are those that translate between the multiplex code and start-stop. Where simplicity of equipment and ease of maintenance are factors of importance, the use of translators could be made unnecessary by 6-unit operation, with start pulses transmitted, at the same rate of modulation as one of the start-stop rates. One method of doing this is shown in Figs. 2(A) and 2(B). Experience with this type of multiplex indicates that it is most reliable, with minimal maintenance problems, and that it is popular with technicians.

As the switching operations of the circuits of Figs. 2(A) and 2(B) are relatively simple, the electronic circuitry required for replacing the rotary distributor would be relatively simple. Ring counters or reliable decade switching tubes are obvious possibilities.

SYNCHRONISM

The heart of any synchronous system is the method used for maintaining optimum time relationship of the receiver with the incoming signals. Any synchronous operation on the part of private networks will probably be almost entirely point-to-point with nothing resembling the complex chains set up by the communication companies. Therefore synchronizing problems should be simpler.

At receiving points, the use of highly accurate but complex time bases would probably be undesirable because of expense, complexity, lack of range, and sluggishness in falling into phase. A relatively simple, inexpensive, and flexible arrangement would be preferable.

At the sending end, a reasonable degree of accuracy and stability would be required; say, an accuracy of one part in 500 and a stability of one part in 10,000. At the receiving end, there should be a range of about 1% for setting the basic frequency, and the synchronizer should be able to maintain optimum time relationship with a change in relative speeds of about one part in 2,000. If signals are cut off, the receiver should not shift phase significantly over a period of a few minutes.

It is probable that methods now in use by the communication companies could be adapted to meet these requirements.

PERFORMANCE SPECIFICATIONS

To meet the needs of typical private telegraph networks, performance specifications should include the following points:

1. Circuitry should be kept reasonably simple so that it can be understood by available technicians.
2. Equipment should be capable of running for some months without servicing; maintenance should not require the services of highly skilled specialists.
3. Where suitable, for the sake of simplicity and ease of maintenance, the modulation rate should be at one of the start-stop rates, that is, 45.45, 56.88, or 74.2 bauds, or some multiple of one of these rates for multichannel circuits.
4. Sending-end frequency accuracy should

be one part in 500 and over-all stability one part in 10,000.

5. At the receiving end, frequency should be adjustable over a minimum range of 1% and the synchronizing circuitry capable of maintaining a constant selection phase with a relative speed drift of one part in 2,000.

6. For multichannel equipment, channel phasing should be semiautomatic on idler signals and in unit element steps on mixed signals.

Conclusions

Private networks might gain significant advantages from the use of synchronous equipment under the following circumstances:

1. Where higher speed trunk circuits are required.
2. Where extra channels are required but are not otherwise available, or costs are high.
3. Where circuits are unstable because of factors such as variable amplitude, interference, and excessive distortion, and where the error rate is high with normal start-stop operation.

It is doubtful if private networks, with some exceptions, could make effective use of the highly complex and expensive synchronous multiplex equipment in general use by the communication companies and the armed forces. For general use in private networks, equipment should be relatively simple and straightforward, easy to maintain, and the price should be equitable.

Although reference was made in this paper to synchronous methods used by the various communication companies, it was not possible to supply much in the way of specific information on such methods. Those interested in a comparison of multiplex techniques may wish to consult the list of references.

Appendix I. Distortion

Telegraph Distortion may be defined as the deviation in the timing of a received transition, after signals have passed through a limiter, relay, flip-flop, or saturated grid, as the case may be, from that of an ideal signal.

Any one transition can have but one value of distortion. Hence, to have meaning, telegraph distortion must be observed over a period of time and expressed as the ratio of the sum of maximum negative and positive deviations observed against the time of a unit element. For example, if the time of a unit element was 20 milliseconds, and the maximum transition variations observed were 2 milliseconds early and 2 milliseconds late, with reference to the center of the transition band, the distortion would be 4/20, or 20%.

On synchronous circuits, the timing reference may be derived from the instrumental time base, which is normally synchronized by an integration of received transitions, or from a stable independent oscillator. On start-stop circuits, the transitions of the selecting pulses, and front transition of the stop pulse, are matched against the beginning of the start pulse.

Sources of Telegraph Distortion

Characteristic Distortion. This is a function of the transmission medium. It can be observed directly only before a signal passes through a limiter and has several facets which include:

1. Asymmetric distortion which is indicated by a transition having a different shape when going from mark to space than when going from space to mark. It is caused by nonlinearity in amplifiers and inductances. Asymmetry is difficult to avoid on any signal that is passed through an amplifier.
2. Short sways which may be derived from the transient components of the signals, or from a tendency for flats to build up or fall away. This effect frequently results in the first element after a flat being too thin or too fat.
3. An exponential rate of change on transitions between mark and space which may be relatively fast on a wide-band circuit and slow on a narrow-band one. Susceptance to sways, asymmetric distortion, interference, and biases is normally inversely proportional to this rate of change and it is probably the most important factor in determining circuit margin.
4. Phase distortion which may be manifested by an overshoot on a transition, or by "broken-back" double and triple element flats.

Interference. On physical lines, interference from power circuits is usually the worst offender, with cross fire from adjacent telegraph circuits a close second. Interference of this type is most frequently developed in common earth impedances. On voice-frequency carriers, cross fire from adjacent channels or other circuits may often be observed. Sometimes interference from power mains is significant. On radio circuits, atmospherics and cross fire from other circuits are usually the dominant sources.

Transitory Biases. These may be due to variation in signal amplitude or by voltage variations in amplifiers.

Keying Losses. The amount of telegraph distortion from this source will depend on the value of n per $2f$, where n is the rate of speed in bauds and f is the carrier frequency in cps. If the value of n is 100 bauds, and f 500 cycles, the keying loss would be 10%. However, if f were 5,000 cps, the keying loss would be only 1%. With amplitude modulation, keying losses can be reduced to zero by synchronizing signals and carrier frequency.

Clipping. Where relays are used in "on-off," or nonpolar coupling, some clipping is difficult to avoid on either mark or space because of relay transit time. The effect may be caused by relay transit time being either under or over compensated.

Additional factors, equivalent to distortion, but applicable to start-stop only:

Motor Differential. When the respective drive motors of the sending and receiving machines are synchronous, and the driving power is derived from independent supply systems, there will be some difference between rate of modulation and the rate at which the receiving printer selects. Where governed motors are used, this is not an uncommon source of trouble.

Instrumental Maladjustments. On the sending side, distortion may occur from an inaccurate cam assembly or from a poor brush condition on a face-plate. On the receiving side, it is rare to find optimum adjustments on a start-stop printer.

On local circuits, there is no significant distortion and on such service the start-stop receiver is ideal, giving minimum trouble with a very low error rate. On trunk circuits, all of the listed factors may be present and, for start-stop operation, margin has to be provided for motor differential, instrumental maladjustments, clipping, and keying losses; this leaves little margin for adverse transmission conditions which include transitory biases, interference, and characteristic distortion. Hence, the error rate, with a start-stop receiver, is apt to be high on long-distance circuits.

On the other hand, with a synchronous system there is no problem of speed differential, and instrumental maladjustments and clipping will be insignificant if polar couplings are used throughout. Hence, practically all of the margin of a circuit operated synchronously can be used to override characteristic distortion, interference, and keying losses. On a long-distance circuit, the error rate should be insignificant for as much as 40% distortion.

References

1. AN IMPROVED 6-CHANNEL TIME DIVISION MULTIPLEX FOR SUBMARINE CABLE TELEGRAPHY, H. F. Wilder. *AIEE Transactions*, vol. 77, pt. I, May 1958, pp. 217–22.

2. ATLANTIC TELEGRAPH CABLE CENTENNIAL, H. H. Haglund. *Western Union Technical Review*, New York, N. Y., vol. 12, no. 3, pp. 81–101.

3. A TRANSISTORIZED TIME-DIVISION MULTIPLEX TELEGRAPH SET, F. D. Biggam. *AIEE Transactions*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 780–85.

4. MULTIPLEX TELEGRAPH EQUIPMENT FOR RADIO AND SUBMARINE CABLE CIRCUITS, B. S. Diamond. *Ibid.*, vol. 76, pt. I, May 1957, pp. 175–82.

◆

Discussion

R. S. Wishart (Western Union Telegraph Company, New York, N. Y.): For 45 years Western Union has been a great user of synchronous methods of operation and none of us will question Mr. Diamond's thesis that such methods have an advantage over start-stop in the face of disturbance that will interfere with operation of the latter. We also use synchronous methods on cables, where usable frequency spectrum is limited and the maximum capacity must be obtained. High-speed data circuits are another case where synchronous methods may be used.

Nevertheless, there are many good reasons for using start-stop operation where plant facilities are relatively free of disturbance and constantly under observation for control of interference and bias.

The paper did not indicate whether the comments were directed at wire and radio telegraph networks owned by individual corporations such as pipe lines, railroads, etc., or whether the references were to private wire systems furnished to others by Western Union and American Telephone and Telegraph Company. Several of the statements supporting synchronous application, therefore, require discussion. Some of them were evidently based on conditions prior to the recent increased availability of 100-wpm teleprinter equipment and facilities to go with it.

If we consider the application of synchronous methods to the Western Union private-wire systems, we are faced with certain problems. More than half of our leased facilities are used as part of switching systems and a considerable portion of the remainder are way station circuits at various speeds and using "hub" type repeaters. A majority in number of the switching systems are of a single switching center type with long way circuits extending from the switching center. Even where there are multiple switching centers, these individual centers serve their areas with way circuits. These way circuits are duplex circuits carrying control signals in both directions. These carefully planned switching and way wire systems are not susceptible of operation by synchronous means in the present state of the art.

There is a constant tendency to increase the speed of way circuits and trunks from 65 to 75, and then to 100 wpm as load increases; contrary to Mr. Diamond's statement it is not difficult to obtain 100-wpm circuits. There is often economy in the higher speed because more stations may be added to a way circuit without greatly increasing the wire mileage paid for. In addition, if one way circuit has been cancelled by shifting its stations to the one with increased speed, positions in the switching center are reduced.

Our 65-, 75-, and 100-wpm circuits are obtained from 150-cycle carrier channels. This provides flexibility compared with the fixed speeds of any particular synchronous system.

Mr. Diamond states that when trunk circuits are lengthy, and comprise two or more sections in tandem, the error rate may be quite high and that in one normally complex switching network, observations indicated an error rate of one per 1,000 characters. He also observes that in many error checks on complex chains comprising start-stop and synchronous links, the majority of errors occurred on the start-stop links.

This condition does not register with me as fitting a Western Union situation. In the first place, we have determined that on our carrier channels and short single wire extensions the following are safe limits on numbers of sections without regeneration:

Speed in Wpm	No. of Sections
65	6
75	5
100	3–4

One or more of these sections may be a physical wire. Channels derived from radio beam systems are included in this calculation, and we can regenerate new signals and continue the expansion to any size network we desire.

As to accuracy, notwithstanding many over-all and sectional tests in our public message and customer's private wire systems, we cannot substantiate an average error rate as bad as one per 1,000 characters. Comparing originating tapes, relay tapes at switching centers, and terminal copies, in a typical large customers' switching system, having several switching centers, we found that the various links each produced 99+% message accuracy and that the origin-to-destination percentage of perfect messages transmissionwise was 96%. In this study, corresponding originating and terminating tapes were compared with each other by means of comparing transmitters and page copies were made from original tapes and compared with terminal copies. A very exact listing of errors incurred during transmission was obtained. Few of the errors could be attributed to the transmission medium. As a result of analysis, equipment faults in the transmitters, perforators, and reperforators were considered to be the cause of most errors.

As engineers, we are not satisfied with this degree of accuracy and don't boast about it, but in switching systems there is at present considerable difficulty in quickly catching a source of error.

As you know under the older systems of operation, complete circuits were operated from terminal to terminal and messages were subject to constant scrutiny at the terminal points. The occasional facility or equipment difficulties which occurred even on the highest class circuits with good maintenance came to light very quickly and were remedied. These same good circuits, when assembled into a switching system, are likely to produce many more of the occasional type error because the switching operators and the mechanical devices now only scrutinize the routing section and sequence numbers of messages. Under such conditions the errored messages become scattered as they are switched and there is no concentrated scrutiny of the messages on each section to catch a minor failure. Such minor failures might be a sticking pin in a transmitter or sticky operation of a selector bar in a reperforator, or a momentary fault in a facility for instance. If error-checking devices can be economically inserted in such systems the error-frequency rate will be drastically reduced.

When we look at records of circuits it is remarkable how many of them are error-free for long periods of time. With our rapid changeover to voice-frequency channels and paired facilities, the Western Union system is now only slightly effected with disturbance from power mains. It is quite different from the old days.

Mr. Diamond comments that transmission channels derived from radio beams are sometimes so seriously afflicted with atmospherics and multipath effects that communication with normal start-stop methods of operation is submarginal. This comment may be accurate as regards some types of radio beam circuits but it does not agree with our findings on Western Union beam circuits. They were specifically designed for telegraph operation with space diversity reception in the older installations and frequency diversity in the newer installations. While there are some residual multipath and fading effects during periods in the summer, they occur after midnight and have cleared by 8 a.m. and have little effect during general business hours. Our 12 years' experience in operating these radio beam circuits indicates that very little more work is needed to reduce the fading troubles to a negligible value. We are reaching a point where equipment outages and "hits" on intracity and office cables outweigh propagation failures.

We don't question the value of synchronous methods in certain difficult transmission conditions or where desired capacity can be obtained in no other way, and we will certainly use that tool when it is required. Actually some 25% of our channels between switching centers in the public message system are derived from synchronous multiplexes operating over 150-cycle carrier channels. This is a rugged, partly electronic, synchronous multiplex designed for the purpose of connecting to start-stop printers and gives good service. We have studied the possibility of expanding this use by developing entirely electronic systems, but the economics of obtaining channels this way as against more 150-cycle carrier channels have not proved satisfactory. The field for increased use of synchronous multiplex in the public message system is not large, when we consider the desirability of distributing channels between cities into alternate routes for protection of service.

Our Western Union facilities are now so stable that we care very little whether synchronous or start-stop operation is applied to the derived telegraph channels. The results are equally good.

B. S. Diamond: Mr. Wishart's comments are very informative and the only point on which there seems to be any serious difference of opinion between us is on error rates.

Error rates on circuits operated by public telegraph systems are unquestionably low. However, support for my observations that it is high in, at least, some private networks is provided by a passage in the *Proceedings* of the Association of American Railroads, May 1958, page 258, which reads:

> Messages normally consist of words which are interpreted by people. Errors in spelling or spacing are often obvious and result in no loss of intelligence. Even when figures are involved, the intelligence of the message may not necessarily be impaired by errors. Because of human ability to judge and interpolate, error rates as high as 1 in 100 are *often* economically acceptable on message circuits.

A Generalized Computer Procedure for the Design of Optimum Systems—Part I

R. RODERICK BROWN
ASSOCIATE MEMBER AIEE

CONVENTIONAL design techniques cannot readily produce optimum solutions to present-day design problems. The ideal characteristics of a design procedure are reviewed herein and compared with the conventional techniques. The characteristics of a currently feasible procedure are noted, and a new technique for use with a digital computer is described. A sample design optimization is employed to illustrate the operation of the new technique, which closely approaches the ideal design procedure.

Characteristics of a Design Procedure

The objective of any system design procedure is usually twofold. First, a design which meets the specifications and other constraints imposed upon the system must be produced. Second, this design should concurrently be optimum with respect to some criterion (cost, weight, etc.). The criterion used to define the optimum system will, of course, depend upon the particular application under consideration. For example, the optimum electric generating system for aircraft is usually defined as the lightest-weight system which meets the specifications, i.e., this system is optimum with respect to weight. For clarity in the ensuing discussion, an optimum design will be defined as the design which simultaneously meets the specifications and is optimum with respect to some criterion, and an acceptable design (i.e., a feasible solution) will be defined as any design which meets the specifications.

The characteristics of the ideal design procedure may now be enumerated. In particular, the procedure should: 1. produce the optimum design, 2. be efficient, 3. be versatile, 4. be accurate, and 5. be currently feasible and economical.

The first requirement is clearly desirable whenever the system is to be manufactured in quantity or is to be used in critical applications. The requirement that the procedure be efficient (i.e., produce a design rapidly) is highly important in competitive fields, and is receiving continual stress in both the military and commercial areas. The requirement of versatility is intimately connected with that of efficiency, for a design procedure independent of both system configuration and complexity will necessarily eliminate the need for adapting it to fit the particular problem under consideration. It is the author's viewpoint that the greatest over-all efficiency will be achieved by utilizing a generalized procedure, as opposed to developing special techniques for each genus of problem. It should be realized that this viewpoint is conjectural, and can be substantiated (or disproved) only by an accumulation of considerable practical experience with both approaches. The requirement of accuracy is fundamental, for the design process would be worthless if the results of its application were not closely related to the physically realizable system. Lastly, the requirement that the procedure be currently feasible and economical is essential if it is to be both immediately useful and practical.

It will become clear from the ensuing discussion that it is presently impossible to realize the ideal procedure. However, it is possible to realize a technique which closely approaches it. Such a technique is described herein. To provide a basis for comparison, it is worth while to describe the conventional design technique and compare it with the ideal.

The Conventional Design Procedure

The conventional procedure for producing an optimum system design is a trial-and-error process involving two overlapping steps. First, an acceptable design is produced. The trial-and-error process consists of choosing certain reasonable values for the design variables (i.e., the physical parameters of the system) and then computing the performance parameters, which are compared with the specifications. Discrepancies between the latter are then eliminated by changing the design variables appropriately. After several variations in the design have been made, the specifications will be satisfied, thus yielding an acceptable design. The second stage of the procedure may now be undertaken. Such additional variations in the design are made that it approaches the optimum more closely. In this manner a series of acceptable designs are made, the best of these being selected as the "optimum."

Clearly, the best of the acceptable designs produced is not necessarily the optimum, or best possible. Indeed, only if an exhaustively large number of designs is produced can there be any confidence that the best available is the optimum. Generally, economic reasons preclude the determination of many designs, with the result that, in practice, the conventional technique does not produce the optimum design.

The choice of the appropriate parameter variations to be made during the design process must, in general, be made by the individual designer. Thus the time required to achieve an acceptable design is determined by the designer's previous experience with and his insight into similar system configurations. Consequently, the procedure will be relatively inefficient when applied to new systems. A further consequence of the need for experience and insight is that the process lacks versatility, for the choice of parameter variations appropriate for one system is unlikely to be appropriate for another. In addition, the design of highly complex systems is usually accomplished by making certain simplifying assumptions, which assuredly increases the efficiency of the technique, but usually also reduces the accuracy of the final results.

It may be noted that current design techniques have capitalized upon the high computational speed of digital computers to effect improvements in the efficiency of the conventional procedure. Indeed, in certain cases, it has been possible to code the computer so that it will automatically produce an acceptable design. Examination of these techniques shows that they fall within the broad category of the conventional procedure and must be considered primarily as implimentations thereof.

It is apparent that the conventional design technique does not fulfill the requirements of the ideal procedure. The

---

Paper 58-234, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 25, 1957; made available for printing November 13, 1958.

R. Roderick Brown is with the Massachusetts Institute of Technology, Cambridge, Mass.

This paper is taken, in part, from an S.M. thesis submitted to the Electrical Engineering Department at Massachusetts Institute of Technology. The work was conducted in the Energy Conversion Group of the Servomechanism Laboratory at the Institute, and was supported, in part, by the Wright Air Development Center under Contract *AF 33(616)3242.*

need for the improved procedure to be described is thus established.

Realization of a Nearly Ideal Procedure

The characteristics common to any realization of a nearly ideal design procedure may now be described. Clearly, the procedure must be placed upon a firm mathematical basis.

The problem of designing an optimum system may be expressed in mathematical form as the problem of minimizing a nonlinear criterion function subject to nonlinear inequality constraints (i.e., as a nonlinear programming problem). Specifically, a problem involving n_t design variables $v(=[v_1, v_2, \ldots, v_{n_t}])$ and m_t functional constraints $g_i(v)$ $[i=1, \ldots, m_t]$ may be expressed as follows: Minimize the criterion function

$$G=G(v) \tag{1}$$

subject to the following constraints:

$$\begin{array}{lll} \text{Type 1:} & g_j(v)=\alpha_j & (j=1, \ldots, m_1) \\ \text{Type 2a:} & g_k(v)\leqslant\alpha_k & (k=1, \ldots, m_{2a}) \\ \text{Type 2b:} & g_p(v)\geqslant\beta_p & (p=1, \ldots, m_{2b}) \\ \text{Type 3:} & \beta_q\leqslant g_q(v)\leqslant\alpha_q & (q=1, \ldots, m_3) \\ \text{Type 4:} & v_r\geqslant\beta_r & (r=1, \ldots, n_4) \\ \text{Type 5:} & \beta_s\leqslant v_s\leqslant\alpha_s & (s=1, \ldots, n_5) \end{array} \tag{2}$$

where

$$m_t=m_1+m_{2a}+m_{2b}+m_3 \tag{3}$$

$$n_4+n_5\leqslant n_t \tag{4}$$

and the upper bounds α and the lower bounds β are nonnegative constants. The design variables v are, of course, the physical parameters of the system and the constraints $g_i(v)$ include not only the performance specifications, but also the constraints necessary to ensure that final design be physically realizable.

It is evident that, in order to guarantee that all of the constraints be satisfied, the design procedure must consider them simultaneously. This aspect of the procedure is in direct contrast to the conventional procedure and to the method of progressive realization,[1,2] both of which consider each constraint individually. These latter methods are at best inefficient, and for many problems may be incapable of providing a solution. In addition, the minimization of the criterion function should, for the sake of efficiency, proceed concurrently with the satisfaction of the constraints. This is also in direct contrast to the aforementioned methods, wherein the minimization is considered as a distinctly separate problem.

At this point it is worth while to recognize that the design problem, in common with other minimization problems, may have several relative minima and that the minimization of the criterion function must be accomplished by employing the methods of the calculus. It will thus be impossible to determine the absolute minimum directly; it can be found only by investigating all possible relative minima. In this sense, it is impossible to realize a technique which will satisfy the first requirement of the ideal design procedure. This is not overly discouraging, however, since the number of relative minima which need be investigated will, in all probability, be small for the majority of design problems. Thus, a procedure which will determine a relative minimum will be nearly ideal and will certainly be a considerable improvement over the conventional technique.

Clearly, any technique which is formulated to solve the optimum design problem must treat the minimization of the criterion function and the satisfaction of the constraints simultaneously, employing the calculus for the former. Techniques previously developed specifically for solving nonlinear programming problems[3,4] do not meet the efficiency and versatility requirements of the ideal design procedure. However, a satisfactory procedure utilizing gradient methods may be formulated.

Design Procedure Using Gradient Techniques

Minimization of Criterion Function

Gradient methods for the minimization of an unconstrained function are well-known;[5] however, they are not directly applicable to a nonlinear programming problem. It is useful to modify the problem statement, equations 1 and 2, such that the inequality constraints are rephrased as equality constraints. This may be done by noting that a constraint of the form $F_i(v)\geqslant 0$ is equivalent to $F_i(v)=z_i^2$ (where z_i is a real variable). The variable z_i is a so-called slack variable, for it absorbs the difference, or "slack," between the value of $F_i(v)$ and 0. With the introduction of slack variables, the problem statement becomes as follows: Minimize

$$G=G(v) \tag{5}$$

subject to the equality constraints

$$f_i(v, z)=0 \quad (i=1, \ldots, m) \tag{6}$$

where

$$z=(z_1, \ldots, z_{m-m_1}) \tag{7}$$

The total number of constraints m is

$$m=m_t+m_3+n_4+2n_5 \tag{8}$$

and the total number of variables n is

$$n=n_t+m-m_1 \tag{9}$$

The classical method for minimizing a function subject to equality constraints is by Lagrangian undetermined multipliers. This method, however, requires that not m but $m+n$ simultaneous nonlinear equations be solved. Consequently, it will not be efficient, nor is it likely to be currently feasible, for the simultaneous solution of such a large number of equations is quite likely to overburden the available computing facilities.

A more feasible method utilizes the fact that the criterion function $G(v)$ is a functional (i.e., a function of functions). The constraints establish an interdependence among the variables, so that m of them are defined in terms of the remaining n-m. The m dependent variables can theoretically be expressed in terms of the n-m $(=k)$ independent variables so that $G(v)$ may in turn be expressed as an unconstrained, explicit

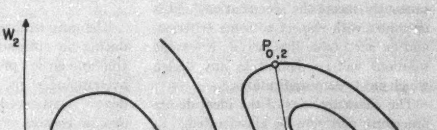


Fig. 1. Method of steepest descent in a space containing two minima

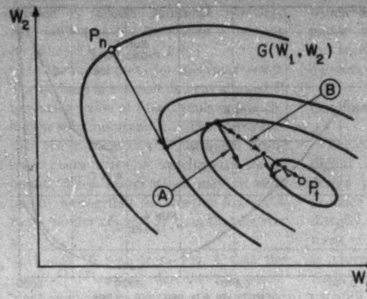


Fig. 2 (left). Optimum gradient method

A—Usual trajectory
B—Accelerated trajectory

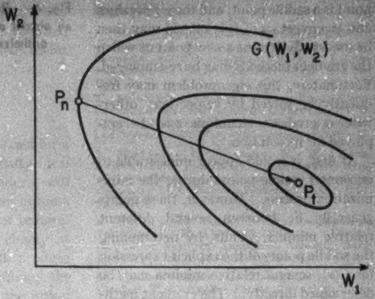


Fig. 3 (right). Newton method

function of the *k* independent variables. Although the explicit expression for $G(v)$ cannot be obtained in practice, it is possible to evaluate its gradient.[1]

The variables *v* and *z* may be considered to be the co-ordinates of an *n*-dimensional space and a particular design may be considered as a point in this space. Within the *k*-dimensional subspace of all acceptable designs there are contour lines of constant $G(v)$, which are, in general, closed lines surrounding the relative minima of $G(v)$. The gradient of the explicit expression for $G(v)$ evaluated at a point may be considered as a vector at that point which points in the direction yielding the largest possible increase in $G(v)$ per infinitesimal unit distance travelled away from the point. The negative of the gradient thus points in the direction of "steepest descent," and is always perpendicular to the contour line through the point in question. Therefore, by moving in the direction of the negative gradient, the criterion function will be decreased as much as is possible for a given infinitesimal distance moved through the subspace.

Finite steps must be taken along the gradient if the minimization is to proceed efficiently. The length of the step must, of course, be controlled so that $G(v)$ does not increase. If the latter is the only criterion used to control the length of the steps, then the gradient method for minimization being used is the familiar method of steepest descent.[6] This method is illustrated in Fig. 1 for a subspace of two dimensions, w_1 and w_2 (i.e., $k=2$). The points $P_{0,1}$ and $P_{0,2}$ are two arbitrarily selected initial points (i.e., initial designs) and $P_{t,1}$ and $P_{t,2}$ are respectively their closest relative minima. It is immediately apparent that the relative minimum which is found by the method is dependent upon the initial point selected, a feature common to all gradient methods. In addition, the method of steepest descent may be subject to time-consuming oscillations. These can be partially eliminated by utilizing the optimum gradient method,[6] wherein the length of the step is adjusted so as to achieve the largest possible decrease in $G(v)$ for the given direction. The oscillations may thus be eliminated in the region far from the relative minimum, while the accelerated optimum gradient method[7] and Newton's method[6] may be used to reduce the oscillations near the minimum; see Figs. 2 and 3, in which the contours of $G(v)$ are very nearly elliptical, indicating that P_n is near P_t.

These four different gradient methods must all be employed during the course of the minimization if the latter is to proceed efficiently. Although these methods have been described previously in the mathematical literature, no criteria for governing their application to the general problem of minimization in the presence of constraints have heretofore been developed. This paper indicates the types of criteria required, and a companion paper describes them in detail.[8]

SATISFACTION OF THE CONSTRAINTS

To evaluate the gradient of the explicit expression for $G(v)$, the *m* constraints must be satisfied at the point in question. Assuming that they are satisfied at a given point (i.e., that the point represents an acceptable design), the gradient methods indicate the proper changes to be made in the *k* independent variables. The *m* dependent variables must then also be changed so as to resatisfy the constraints. These changes will be indicated by the solution of the set of *m* nonlinear constraint equations in *m* unknowns. The most efficient method for solving this problem is the Newton-Raphson method, alternatively known as Newton's method of successive approximations.[9] The efficiency of this method arises because it can be conveniently modified to use information already determined in the computation of the gradient, and because of its potentially rapid convergence. Usually, the changes in the independent variables will be small enough so that the method will rapidly converge. Of course, an additional criterion must be placed on the size of the steps taken so as to ensure the latter. By selecting an appropriate criterion allowing a small error in the constraints, the number of iterations required by the Newton-Raphson method may be reduced to one, thus minimizing the time required to resatisfy the constraints after each change in the independent variables.

The selection of an initial point will be to some extent arbitrary, so that it is unlikely that this point represents an acceptable design. In general, several iterations of the Newton-Raphson method will be required to reduce the error in the constraints to within the allowable tolerance, providing that the method converges. Should the method be divergent, a gradient method may be used to minimize the error. Fortunately, once an acceptable design has been found for the initial point, these lengthy calculations need not be repeated, one iteration of the Newton-Raphson method being sufficient thereafter.

TERMINATION OF THE MINIMIZATION

The gradient methods will terminate at a stationary point, i.e., at a point where the gradient of the explicit expression for $G(v)$ is zero. It is usually necessary to determine if this point is a relative minimum (i.e., a solution) or a saddle point. If it is a saddle point it is necessary to employ a nongradient method to move away from the point, after which the minimization may continue as before.

The stationary point may be tested by examining a matrix which is employed in Newton's method for minimization. If the matrix is not positive definite, the point is a saddle point, and the eigenvalues and eigenvectors of this matrix may then be used to determine a new point at which the gradient methods may be re-employed. Fortunately, this eigenproblem may frequently be solved by inspection; otherwise, a gradient technique may be employed for its solution.

At first, it would appear undesirable to encounter saddle points during the minimization process. However, these points generally lie between several different relative minima. Thus, by determining the saddle points of the explicit expression for $G(v)$, several relative minima may be determined directly. The gradient methods may be constrained to converge to some of the saddle points by an appropriate selection of the initial values of the slack variables. In this fashion, the gradient methods approach the first requirement of the ideal design procedure far more closely than would otherwise be expected.

Utilization of a Digital Computer

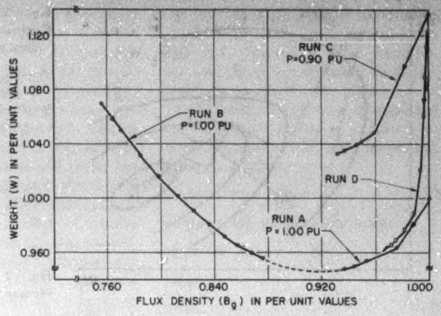
The calculations required by the gradient methods are highly complex and far more detailed than those required by the conventional design procedure. A high-speed digital computer must therefore be employed if the proposed design procedure is to be efficient and economical. The use of a digital computer is currently feasible, as evidenced by its recent application to design problems.[2,10]

At this point, it is worth while to enumerate the characteristics of the ideal computer routine. Specifically, the routine should: 1. be efficient in computer hours and in man-hours, 2. be highly generalized, and 3. require a minimum of manual intervention.

Only by meeting these requirements can the capabilities of the computer be fully exploited. The computer routines currently used by industry for the design of optimum systems are based upon the conventional design procedure and consequently fail to meet these requirements. The gradient methods, however, may be programmed for the computer so as to meet all of them.

Efficient iterative methods for performing the calculations for the gradient methods are available so that a routine requiring a minimum computer time expenditure may be written. The number of man-hours required to program the routine for any particular problem may be minimized by dividing the routine into two sections, one for the basic calculations (partial differentiation, matrix inversion, etc.) and one for the design equations $G(v)$ and the $g_i(v)$. The routine will thus be highly generalized, for it may be used to solve any design problem expressed in the form given by equations 1 and 2. Indeed, the gradient methods may be used to solve almost any nonlinear programming problem. Furthermore, the methods can produce a solution directly from an initial design by a well-defined, iterative process which can be completely automatized. Consequently, no manual intervention in the routine is required to produce a relative minimum. This is in direct contrast to the currently employed routines, wherein intervention is required after the production of each acceptable design.

Fig. 4. Results of an aircraft alternator optimization



An additional advantage in using a digital computer is the high accuracy which can be achieved. The relatively large number of significant figures which may be carried throughout the calculation coupled with reasonably accurate approximation equations will thus ensure that the fourth requirement of the ideal design procedure is met.

Example of a Design Optimization

A computer routine has been written to illustrate the applicability of the gradient methods to system optimization. For the purpose of simplification, a routine using only the method of steepest descent was coded for the Whirlwind I computer at the Massachusetts Institute of Technology. The system chosen for study was a typical 40-kva 120/208-volt 3-phase 400-cycle 6,000-rpm aircraft alternator, which was to be optimized with respect to weight W. The machine was implicitly constrained to meet these specified ratings. The total full-load, zero-power factor losses P were assumed to be the only explicit constraint $g_i(v)$ which need be considered during the optimization. The design variables chosen were the armature stack length L and the air-gap flux density B_g. All other physical parameters were held fixed.

The results of computer runs on four different problems are presented in Fig. 4. A commercial design was chosen as the base for the pu (per-unit) values of W, P L, and B_g. This design was expected to be close to the optimum. The data on each intermediate point corresponding to an acceptable design were printed by the computer in order to provide a basis for evaluating the efficiency of the method of steepest descent. Notice from Fig. 4 that B_g was found to be an independent variable. All runs were stopped when the final design differed from the optimum by no more than one per cent. This point was determined approximately, as the criterion for detecting the minimum had not been coded.

For runs A and B, the design problem was expressed as follows: Minimize the criterion function $W=G(L,B_g)$, subject to the equality constraint (type 1) $P=g(L, B_g)=1.0$.

With the commercial machine as an initial design, run A took only four steps to reach a solution. To verify this solution, run B was made from an initial design for which $B_g=0.75$ pu and L was adjusted to ensure that $P=1.0$ pu. Because this initial point was further from the solution, ten more steps were required than for run A.

For run C, the design problem was expressed as follows: Minimize $W=G(L, B_g)$ subject to the equality constraint (type 1) $P=g(L,B_g)=0.9$.

The constraint on P was not exactly satisfied at the initial point. However, the Newton-Raphson method rapidly reduced the error in P, thus indicating that small initial errors are tolerable, for they will be reduced during the optimization process.

For run D, the design problem was ex-

Vol. 7
Pt. 1
1959

pressed as follows: Minimize $W=G(L, B_g)$, subject to the inequality constraint (type 2a) $P=g(L,B_g) \leqslant 1.0$.

At the initial point for this run $P \approx 0.9$. The method of steepest descent allowed P to reach a value of 1.0 pu while approaching the solution to run A. This illustrates that the solution to this particular problem occurs when P reaches its upper bound, a conclusion which may be verified by noting that the solution for run A has a smaller value of W than that for run C.

It may be seen from Fig. 4 that the method of steepest descent was somewhat inefficient in producing a solution to run B, for which $k=m\text{-}n=1$, and particularly so for run D, for which $k=2$. The use of the other gradient methods, particularly the optimum gradient method, would have improved the efficiency greatly. However, these runs should not be evaluated on the basis of their efficiency, which should be expected to be relatively low because of the restricted use of the gradient methods. Rather, their value lies in illustrating the feasibility of using these methods for the design of optimum systems. It may be noted that while Whirlwind I required approximately 20 seconds to determine each step,[11] the use of a computer such as the International Business Machines Corporation *704* coupled with the improved calculational techniques which are described in the companion paper[8] would reduce the time required for each step of this problem to about one second.

Conclusions

The gradient methods when coded for a digital computer represent a nearly ideal design procedure. This procedure will produce one or more relative minima directly, thus satisfying the first requirement of the ideal procedure as closely as is currently possible. It will produce these solutions rapidly by utilizing the capabilities of the computer to the fullest extent, thus meeting the second requirement. The third requirement is also satisfied because the procedure is specifically designed to solve the general design problem. The fourth and fifth requirements are met by using a digital computer to perform the necessary calculations.

This paper has outlined the general philosophy of design which has led to the development of the gradient design procedure, and has described its general characteristics. In essence, the procedure consists of a complex group of criteria which have been developed to unite a number of previously known mathematical techniques. These criteria, which represent the core of the method, are described in detail in reference 8.

References

1. System Synthesis With the Aid of Digital Computers, **J. B. Dennis, R. F. Nease, R. M. Saunders.** *AIEE Transactions*, vol. 78, pt. I (*Communication and Electronics*), 1959 (*Paper no. 59-196*).

2. Induction Machinery Design Being Revolutionized by the Digital Computer, **C. G. Veinott.** *Ibid.*, vol. 75, pt. III (*Power Apparatus, and Systems*), 1956 (Feb. 1957 section), pp 1509–17

3. Nonlinear Programming, **H. W. Kuhn, A. W. Tucker.** *Proceedings*, Second Berkeley Symposium on Mathematical Statistics and Probability, University of California, Berkeley, Calif., 1951, pp. 481–92.

4. An Analog Solution of Programming Problems, **C. M. Ablow, G. Brigham.** *Journal*, Operations Research Society of America, Baltimore, Md., vol. 3, no. 4, Nov. 1955, pp. 388–94.

5. Gradient Methods of Maximization, **J. B. Crockett, H. Chernoff.** *Pacific Journal of Mathematics*, Tokyo, Japan, vol. 5, 1955, pp. 33–50.

6. Gradient Methods in the Solution of Systems of Linear Equations, **M. L. Stein.** *Journal of Research*, National Bureau of Standards, Washington, D C., vol. 48, 1952, pp. 407–13.

7. Acceleration of the Optimum Gradient Method, Preliminary Report (abstract), **G. E. Forsythe, T. S. Motzkin.** *Bulletin*, American Mathematical Society, Providence, R. I., vol. 57, 1951, pp. 304–05.

8. A Generalized Computer Procedure for the Design of Optimum Systems—Part II, **R. Roderick Brown.** *AIEE Transactions*, pp. 289–93 of this issue.

9. Introduction to Numerical Analysis (book), **F. B. Hildebrand.** McGraw-Hill Book Company, Inc., New York, N. Y., 1956, p. 451.

10. Application of a Digital Computer to the Design of Power Transformers to Specifications, **C. L. Moore, W. T. Duboc, P. A. Zaphyr.** *AIEE Transactions*, vol. 75, pt. I (*Communication and Electronics*), May 1956, pp. 134–38.

11. Design of Systems Using Digital Computers, **R. R. Brown, J. B. Dennis, C. Kingsley.** *WADC Technical Note 56-383*, Wright Air Development Center, Dayton, Ohio, Nov. 1956, pp. 23–28, 41.

A Generalized Computer Procedure for the Design of Optimum Systems—Part II

R. RODERICK BROWN
ASSOCIATE MEMBER AIEE

THE general characteristics of a new and powerful procedure for the design of optimum systems have been outlined in Part I of this paper.[1] This procedure consists of a complex group of criteria specifically formulated to enable previously known gradient minimization techniques to be applied to the design problem.

This part of the paper describes in detail the criteria developed to govern the application of the gradient techniques, and concurrently presents an abbreviated discussion of the latter. In addition, the procedures required to effect a digital computer solution are described.

Preliminary Considerations

The problem of designing an optimum system may be expressed either as a nonlinear programming problem (equations 1 and 2, Part I[1]) or, by introducing the slack variables $\mathbf{z}$, as a problem in minimizing a function subject to equality constraints (equations 5 and 6, Part I). Specifically, the problem statement is: Minimize the criterion function

$$G=G(\mathbf{v}) \tag{1}$$

subject to the constraints

$$f_i(\mathbf{v}, \mathbf{z}) = 0 \quad (i=1, \ldots, m) \tag{2}$$

where

$$\mathbf{v}=(v_1, \ldots, v_{n_t})$$

$$\mathbf{z}=(z_1, \ldots, z_{m-m_1})$$

The total number of constraints m is

$$m=m_t+m_3+n_4+2n_5 \tag{3}$$

and the total number of variables n is

$$n=n_t+m-m_1 \tag{4}$$

The n variables $\mathbf{v}$ and $\mathbf{z}$ must be divided into m dependent variables $\mathbf{x}$ $(=\|x_1, \ldots, x_m\|)$ and $n\text{-}m$ $(=k)$ independent variables $\mathbf{w}$ $(=\|w_1, \ldots, w_k\|)$. It is then

---

Paper 58-235, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 25, 1957; made available for printing November 13, 1958.

R. Roderick Brown is with the Massachusetts Institute of Technology, Cambridge, Mass.

This paper is taken, in part, from an S.M. thesis submitted to the Electrical Engineering Department at Massachusetts Institute of Technology. The work was conducted in the Energy Conversion Group of the Servomechanisms Laboratory at the Institute, and was supported, in part, by Wright Air Development Center under Contract *AF 33(616)3242*.

possible to evaluate the gradient of $G(\mathbf{v})$ of equation 1 by the equation[2]

$$(\nabla_f G)_i = (\mathbf{G}_w)_i - (\mathbf{G}_x)_i \|F_x\|^{-1} \|F_w\| \qquad (5)$$

where the elements of the matrices are as follows:

$\mathbf{G}_w \rightarrow \partial G/\partial w_j \quad (j=1, \ldots, k)$

$\mathbf{G}_x \rightarrow \partial G/\partial x_j \quad (j=1, \ldots, m)$

$\|F_x\| \rightarrow \partial f_i/\partial x_j \quad (i, j=1, \ldots, m)$

$\|F_w\| \rightarrow \partial f_i/\partial w_j \quad (i=1, \ldots, m;\ j=1, \ldots, k)$

provided that the determinant of the matrix $\|F_x\|$, called the Jacobian J, does not equal zero. The variables $\mathbf{x}$ and $\mathbf{w}$ may be chosen from the $\mathbf{v}$ and $\mathbf{z}$ so as to minimize the possibility that $J=0$. It is particularly important to set any slack variable z_i equal to an independent variable w_j if that variable equals zero or is likely to become zero during the solution of the problem. In addition, the restriction

$$\mathbf{G}_z \neq 0 \qquad (6)$$

must be met, so that it is recommended that as many as possible of the slack variables be termed independent variables. Thus, the design variables $\mathbf{v}$ associated with the type 4 and 5 constraints (see Part I, equation 2) will be termed dependent variables, and the necessary restriction

$$m-n=k>0 \qquad (7)$$

becomes

$$n_4+n_5<n_1-m_1 \qquad (8)$$

if these variables are to be considered explicitly during the solution. (Note that most constraints of the form $v_j \geqslant 0$ need not be considered.) The remaining variables should then be chosen so that the diagonal elements of $\|F_x\|$ are as large as possible in comparison with the off-diagonal elements, consistent with meeting the restriction imposed by equation 6. This will contribute greatly to the minimization of errors incurred during the inversion of $\|F_x\|$ by making $\|F_x\|$ as well-conditioned as possible.

Gradient Methods for Minimization

The recurrence equations

$$\mathbf{w}^{p+1} = \mathbf{w}^p + \Delta\mathbf{w}^p \qquad (9)$$

and

$$\mathbf{x}_a^{p+1} = \mathbf{x}^p + \Delta\mathbf{x}_a^p \qquad (10)$$

may be used to generate a sequence of points P_p (defined by $\mathbf{w}^p$ and $\mathbf{x}^p$) starting from some initial point P_0 and terminating at the desired minimum P_t. For the gradient methods

$$\Delta\mathbf{w}^p = -\lambda^p \|B\|^{-1} \nabla_f G^p \qquad (11)$$

and

$$\Delta\mathbf{x}_a^p = -\|F_x\|^{-1} \|F_w\| \Delta\mathbf{w}^p \qquad (12)$$

where λ^p is a positive constant used to control the distance between the points P_p and P_{p+1} (i.e., size of the vector step from P_p to P_{p+1}), $\|B\|$ is a $k \times k$ metric imposed on the space,[3] and $\nabla_f G$ is merely the transpose of $\nabla_f G$, as given by equation 5. Assuming for the moment that the constraints $f_i(\mathbf{v},\mathbf{z})$ are satisfied at P_p, it is clear that the change $\Delta\mathbf{w}^p$ will violate the $f_i(\mathbf{v},\mathbf{z})$, and the subsequent change $\Delta\mathbf{x}_a^p$ will resatisfy them only if they are linear and no roundoff or other calculational errors occurs. Since the latter cannot be expected, it is necessary to reduce this error, which can be done most readily by making the additional change in $\mathbf{x}$

$$\Delta\mathbf{x} = -\|F_x\|^{-1}\mathbf{y} \qquad (13)$$

where the elements of $\mathbf{y}$ are $y_i = f_i(\mathbf{v},\mathbf{z})$. This is equivalent to using one iteration of the familiar Newton-Raphson method. The point found employing $\Delta\mathbf{w}^p$ and $\Delta\mathbf{x}_a^p$ will be termed an approximate point $P_{p+1,a}$ and the point subsequently found from this by employing equation 13 will be termed P_{p+1}.

Generally, it is required that the percentage error in the satisfaction of the constraints be less than some small positive constant c_f. Thus each point P_{p+1} will be required to meet the restriction

$$\phi_{p+1} \leqslant c_f \qquad (14)$$

where

$$\phi \triangleq \max_i \left|\frac{y_i}{\gamma_i}\right| \qquad (15)$$

and γ_i is the bound associated with $g_i(\mathbf{v},\mathbf{z})$, i.e., either α_i and β_i. Furthermore, it is reasonable to assume that if equation 14 is satisfied at $P_{p+1,a}$, i.e., if

$$\phi_{p+1,\,a} \leqslant c_f \qquad (16)$$

then the change in $\mathbf{x}$ given by equation 13 will indeed reduce the error $\mathbf{y}$.

Four of the gradient methods for minimization of an unconstrained function may be extended to the problem at hand. These are the method of steepest descent (MSD), the optimum gradient method (OGM), the accelerated optimum gradient method[4] (AOGM), and Newton's method (NM). All methods employ the restriction

$$G^{p+1} < G^p \qquad (17)$$

in order to guarantee convergence to a minimum and, with the exception of NM, all utilize some simple metric $\|B\|$, of which the identity (or unit) matrix is probably the best for a general design problem. Aside from the restriction that equation 17 be met, the MSD is characterized by an arbitrary selection of the vector step size, i.e., of λ. In contrast, the OGM requires λ to be such that the smallest possible value of G is obtained in the direction given by $\nabla_f G$. The AOGM employs a special equation described later, and is constructed so as to prevent the OGM from settling into a time-wasting periodicity. Lastly, NM employs a special matrix of second derivatives, $\|L^*\|$ for the metric $\|B\|$ in conjunction with either the MSD or the OGM, and it can be guaranteed to converge rapidly only when $\|L^*\| \approx$ constant and is positive definite; otherwise, NM may be divergent.

Because of the possibility of predicting the value of G at P_{p+1} from that at $P_{p+1,a}$, the restriction expressed by equation 17 will be supplemented, but not supplanted, by the restriction

$$G_a^{p+1} < G^p \qquad (18)$$

Selection of a Minimization Technique

For maximum efficiency, the OGM should be employed whenever possible. In the vicinity of the minimum, the AOGM may be used to supplement the OGM, and very close to the minimum, NM may be used to advantage. In those situations wherein the OGM cannot be employed, then of necessity the MSD must be utilized.

Initially, one step in the MSD must be taken, using some arbitrary value for λ. The resultant point $P_{p+1,a}$ thus determined must be such that the restrictions given by equations 16 and 18 are satisfied. If they are not, the value of λ must be changed by multiplying it by a multiplier λ_m:

$$\lambda_m = \begin{cases} 0.5\,\dfrac{c_f}{\phi} \text{ if } \phi \geqslant c_f \\ 1 \text{ if } 0.01\,c_f \leqslant \phi < c_f \text{ and } G_a^{p+1} < G^p \\ 10 \text{ if } \phi < 0.01\,c_f \text{ and } G_a^{p+1} < G^p \\ 0.5 \text{ if } \phi < c_f \text{ and } G_a^{p+1} \geqslant G^p \end{cases} \qquad (19)$$

where ϕ and, of course, G_a^{p+1} are to be evaluated at $P_{p+1,a}$. It may be shown that this formulation for λ_m is stable and will yield a value $\lambda_m = 1$ (at which time $P_{p+1,a}$ is a tentatively acceptable point) after a reasonably short number of iterative calculations. The point P_{p+1} is then found and the restrictions expressed by equations 14 and 17 are checked. If they are not satisfied, then the prediction of the characteristics of P_{p+1} from those of $P_{p+1,a}$ was in error, so that equation 19 must be replaced by

$$\lambda_m = \begin{cases} 0.5 \text{ if } \phi > c_f \text{ or } G^{p+1} \geq G^p \\ 1 \text{ otherwise} \end{cases} \tag{20}$$

where ϕ and, of course, G^{p+1} are evaluated at P_{p+1}.

Having determined P_{p+1} by the MSD, P_{p+2} must be found either by the MSD or the OGM. If neither equation 17 nor 18 has been violated during the attempt to find P_{p+1}, then the OGM may be employed by utilizing $s\mathbf{w}^p$ and $s\Delta\mathbf{x}_a{}^p$ to find $\mathbf{w}^{p+s}$ and $\mathbf{x}_a{}^{p+s}$ $(s=2, 3, \ldots)$, equation 13 being employed after each step. In this fashion, a total of s steps is taken using the gradient evaluated at P_p, $\nabla_f G^p$. This procedure (the OGM) must be terminated as soon as equation 14, 16, 17, or 18 is violated, at which time the MSD is again employed, $\nabla_f G$ being reevaluated at the last acceptable point found. Clearly, a true OGM step has been made only if the method terminates because either equation 17 or 18 is violated.

When a termination occurs because either equation 17 or 18 is violated, it is possible to obtain a better approximation to the "optimum" point in the direction given by $\nabla_f G^p$. Let P_{p+s-1} be the last acceptable point found. Then there are three equally spaced points P_{p+s-2}, P_{p+s-1}, and $P_{p+s,a}$ available, where

$$G^{p+s-2} > G^{p+s-1} < G_a{}^{p+s}$$

The improved approximation to the optimum point, $P_{p+s,a,q}$ may be found from P_{p+s-1} by quadratically interpolating between P_{p+s-2} and $P_{p+s,a}$ thus

$$\mathbf{w}^{p+s,q} = \mathbf{w}^{p+s-1} + \lambda_q \Delta\mathbf{w}^p \tag{21}$$

$$\mathbf{x}_a{}^{p+s,q} = \mathbf{x}^{p+s-1} + \lambda_q \Delta\mathbf{x}_a{}^p \tag{22}$$

where

$$\lambda_q = \frac{G^{p+s-2} - G_a{}^{p+s}}{2(G^{p+s-2} - 2G^{p+s-1} + G_a{}^{p+s})} \tag{23}$$

These equations are directly applicable when equation 18 is violated at $P_{p+s,a}$; if equation 17 is violated at P_{p+s}, then the foregoing equations also apply after striking the subscript a from equation 23. If

$$G^{p+s,q} < G^{p+s-1} \tag{24}$$

then $P_{p+s,q}$ is the desired optimum point. If equation 24 is not satisfied, then no optimum point can be found, and P_{p+s-1} must be used as the launching point for a step in the MSD.

The distinguishing characteristic of the OGM as described herein is that the gradient $\nabla_f G$ is not evaluated at every point P_p, in contrast to the MSD, where it is. This distinction is important, since an optimum point may be found in the MSD if λ is reduced because either equation 17 or 18 is not satisfied. A point $P_{p+1,a,q}$ (or $P_{p+1,q}$) may be found from P_p in a manner similar to that described for the OGM. Given the three equally spaced points P_p, $P_{p+1,a*}$, and $P_{p+1,a**}$, where

$$G_{a*}{}^{p+1} < G^p < G_{a**}{}^{p+1}$$

the appropriate λ_q is

$$\lambda_q = \frac{G_{a**}{}^{p+1} - 4G_{a*}{}^{p+1} + 3G^p}{2(G_{a**}{}^{p+1} - 2G_{a*}{}^{p+1} + G^p)} \tag{25}$$

which is used in conjunction with equations 21 and 22 $(s=1)$.

The AOGM (accelerated OGM) may be used only when the metric $||B||$ is the unit matrix, when two or more optimum points have been found in succession, and when the current point P_p is near the minimum. Considering the three points P_a, $P_b(=P_{a+s_a,q})$, and $P_p(=P_{b+s_b,q})$ and defining

$$V \triangleq \begin{bmatrix} \mathbf{x} \\ \cdots \\ \mathbf{w} \end{bmatrix} \tag{26}$$

the AOGM utilizes the equation

$$\mathbf{V}^{p+1,a} = \mathbf{V}^p + \frac{1}{s_a + s_b}(\mathbf{V}^p - \mathbf{V}^a) \tag{27}$$

to determine the point $P_{p+1,a}$. It will be said that P_p is sufficiently close to the solution to allow the AOGM to be applied if

$$\mu < 4c_v \tag{28}$$

(μ and c_v are defined in the following section.)

The application of NM is described in the following section. It may be noted in passing that an optimum step will be taken in NM if $\lambda = 1$.

Detection of a Solution

The gradient methods will, perforce, terminate at a point where the gradient is zero, i.e., at P_t

$$\nabla_f G^t = 0 \tag{29}$$

This stationary point P_t may be either a relative minimum or a saddle point (it cannot be a relative maximum because of the restriction that equation 17 be satisfied). Now, a relative minimum is essentially a stable equilibrium point, for it is completely surrounded by regions having a higher value of G, whereas a saddle point is an unstable equilibrium point, there being regions in its vicinity which have a lower value of G. Consequently, the gradient methods (the MSD and the OGM) will generally converge to a relative minimum except when they are specifically constrained to converge to a saddle point. Since it is quite likely that there are several relative minima present in a design problem, it is clearly desirable to locate several such minima, so that the best of these may be chosen for the final optimum design. Clearly, if all such minima were located, then the best of them would be the true optimum (i.e., the absolute minimum). Since there may be several relative minima in the vicinity of a saddle point, it is thus desirable to locate saddle points.

It is known that if the functions $g_i(\mathbf{v})$ are convex (i.e., have positive second partial derivatives) for positive values of the variables $\mathbf{v}$, the solution (a minimum) will occur when all constraints are equal to their bounds,[5] i.e., when the independent slack variables $\mathbf{z}'$ are zero. Recognizing that for many design problems the $g_i(\mathbf{v})$ are not convex everywhere, it is apparent that if $\mathbf{z}'=0$ at P_t, then it is quite likely that P_t is a saddle point rather than a relative minimum. Now, if at the initial point P_0 $\mathbf{z}'=0$, then, throughout the minimization $\mathbf{z}'$ will remain equal to zero, so that these k_z variables $\mathbf{z}'$ may be eliminated from the problem (thus reducing its size) until P_t is reached, at which point they are reintroduced.

The resultant point P_t must then be tested to determine if it is a saddle point and, if so, a method for continuing the solution must be described. Consider the symmetric matrix of second partial derivatives $||L^*||$, having the elements $\partial^2 G_e/\partial w_i \partial w_j$ $(i, j=1, \ldots, k)$, where $G_e = G_e(\mathbf{w})$ is the explicit expression for $G(\mathbf{v})$ (equation 1) which could theoretically be obtained by solving the constraints (equation 2) explicitly for $\mathbf{x}$, and then substituting $\mathbf{x}$ into equation 1, thus yielding an expression in terms of $\mathbf{w}$ alone (note that $\nabla_f G \equiv$ gradient G_e). If $||L^*||$ is positive definite at P_t, then this point is a relative minimum. On the other hand, if $||L^*||^t$ is semidefinite, then P_t is a saddle point, and the appropriate directions of descent from P_t are given by the eigenvectors associated with the negative eigenvalues of $||L^*||$. For $\mathbf{z}'=0$ at P_t it is reasonably safe to say that the only negative eigenvalues of $||L^*||$ are the negative $\partial^2 G_e/\partial z'_j{}^2$, in which case the appropriate eigenvector is simply $\mathbf{e}_j$ ($\mathbf{e}_j$ is a column matrix, the only nonzero element of which is the jth element, which is unity). Thus, when $\mathbf{z}'=0$, it is necessary to examine only the $\partial^2 G_e/\partial z'_j{}^2$ at P_t. Of course, if these are all positive, P_t is a relative minimum. These partials may be calculated most readily by the numerical procedure described in the following section.

Usually it is not necessary to determine the minimum P_t exactly. An approximate solution $P_{t,a}$ is generally acceptable if it can be guaranteed to be within a certain specified distance of P_t. Manufacturing specifications commonly place a tolerance on the values of the variables $\mathbf{v}$ which can be achieved during fabrication of the system. This tolerance may be expressed as a certain percentage of the value of the $\mathbf{v}$, say, c_v. It may be expected that G will vary inappreciably over a region about P_t so that any physically realizable design determined by 1.5 c_v will be acceptable. In this case, it can be shown that if the direction of a vector step is towards P_t, an acceptable solution $P_{t,\alpha}$ will be found when

$$\mu \leqslant c_v \tag{30}$$

where

$$\mu \triangleq \max_j \left| \frac{\Delta V_j}{V_j} \right| \tag{31}$$

ΔV_j being the change in V_j determined by $\Delta \mathbf{w}^p$ and $\Delta \mathbf{x}^p$. Thus, if equation 30 is satisfied during a step in the AOGM, this method should be abandoned in favor of the MSD and, once equation 30 is again satisfied, then the MSD must be abandoned in favor of NM. Of course, if the zero-valued $\mathbf{z}'$ variables have been eliminated from the problem, then the relevant elements of $\|L^*\|$ must be eliminated, leaving a reduced matrix $\|L_{w'w'}\|$, which, by previous assumption, is positive definite, thus guaranteeing that NM would converge to P_t. Once equation 30 is satisfied in NM, $P_{t,\alpha}$ has been determined. The saddle point test described previously may be applied at $P_{t,\alpha}$, for it is assumed that $\|L^*\|^{t,\alpha} \approx \|L^*\|^t \approx \text{constant} \neq 0$. It will be recognized that the latter assumption, which is valid for a large class of system problems, is equivalent to assuming second-order behavior over a small region near the minimum.

Digital Computations

The calculations required for the gradient methods (differentiation, matrix inversion, etc.) can best be performed by numerical techniques which, fortunately, are well suited for digital computation. Indeed, if a digital computer is not employed, the time required to solve the design problem would be prohibitively great.

A finite difference equation must be employed to evaluate the first partial derivatives required for the matrices $(\mathbf{G}_z)_t$, $(\mathbf{G}_w)_t$, $\|F_x\|$, and $\|F_w\|$. The best equation which can be employed is that which requires the least computational time. (It is important to realize that it is impossible to specify which equation will give the least error.) This equation is

$$\frac{\partial f_i}{\partial v_j} \approx \frac{1}{h_j}[E_j f_i(\mathbf{v}) - f_i(\mathbf{v})]$$

$$(i=0, 1, \ldots, m;\ j=1, \ldots, n_t) \tag{32}$$

where E_j^r is a shifting operator which adds rh_j to v_j, and $f_0(\mathbf{v}) \triangleq G(\mathbf{v})$. The spacing h_j may reasonably be made:

$$h_j = \eta v_j \tag{33}$$

where η is determined by the size of the region over which the designer expects linear approximations to be valid (most probably $0.01 \leqslant \eta \leqslant 0.1$). Notice that equation 32 is not used to calculate the partial derivatives with respect to the slack variables $\mathbf{z}$. These derivatives are readily determined analytically; specifically, the only such partials which are nonzero are the

$$\frac{\partial f_i}{\partial z_i} = -2z_i \tag{34}$$

where the f_i are the type 2 through 5 constraints.

Assuming that the explicit expression for the criterion function $G_e(\mathbf{w})$ is available, the elements of the symmetric matrix $\|L^*\|$ may be calculated from

$$\frac{\partial G_e^2}{\partial w_j^2} \approx \frac{1}{h_j^2}[E_j^{-1}G_e(\mathbf{w}) - 2G_e(\mathbf{w}) + E_j G_e(\mathbf{w})]$$

$$(j=1, \ldots, k) \tag{35}$$

and

$$\frac{\partial^2 G_e}{\partial w_j \partial w_i} \equiv \frac{\partial^2 G_e}{\partial w_i \partial w_j} \approx \frac{1}{h_i h_j}[G_e(\mathbf{w}) - E_i G_e(\mathbf{w}) - E_j G_e(\mathbf{w}) + E_i E_j G_e(\mathbf{w})]$$

$$(i<j), \text{ i.e., } (i=1, \ldots, k;\ j=i+1, \ldots, k) \tag{36}$$

The expression for $G_e(\mathbf{w})$ need not be available, however, for

$$E_j^r G_e(\mathbf{w}) = E_j^r G(\mathbf{x}^*, \mathbf{w}) \tag{37}$$

where

$$\mathbf{x}^* \triangleq \mathbf{x}^p - \|F_x\|^{-1}\mathbf{y}^* \tag{38}$$

and the elements of $\mathbf{y}^*$ are

$$y_i^* \triangleq E_j^r f_i(\mathbf{x}, \mathbf{w}) \tag{39}$$

The quantity $E_i^r E_j^r G^e(\mathbf{w})$ may be evaluated similarly.

It has been tacitly assumed previously that the functions $G(\mathbf{v})$ and $g_i(\mathbf{v})$ were analytic functions of the continuous variables $\mathbf{v}$. However, when the differentiation of these functions is performed numerically, this restriction may be relaxed. In particular it is required only that the functions and the variables be continuous.

The matrix calculations may be readily carried out on a digital computer. Perhaps the most economical method for inverting a matrix is the noniterative direct-pivotal Crout method. The calculations required by equations 5, 12, and 13 may be arranged to minimize the calculational time by examining the number of multiplications required. Two basic calculational schemes may be used, the first (method A) being characterized by the immediate evaluation of the inverse of $\|F_x\|$, and the second (method B) by the evaluation of the required inverse products, such as $\|F_x\|^{-1}\|F_w\|$. The relative efficiency of these methods based on the number of multiplications involved depends upon m, k, and s (s being the number of steps taken for each evaluation of the gradient $\nabla_f G^p$). It may be shown that for $m \geqslant 6$, $s \geqslant 2$, method A is more economical on this basis. Furthermore, if $s>1$, this method may require less computer time expenditure in the transfer of information between primary (fast-access) and secondary (slow-access) memories during the operations of the Crout method, which provides additional support in favor of method A.

Errors in the evaluation of the partial derivatives and in the matrix calculations may cause the computed value of the gradient to point in the direction of increasing, rather than decreasing, $G(\mathbf{v})$. It may be said that the error in $\nabla_f G$ is not excessive as long as equation 30 is not met. Once this equation is satisfied, it is desirable to reselect the dependent variables $\mathbf{x}$ in order to ensure that the matrix $\|F_x\|$ is as well conditioned as possible. A generalized analysis of the errors propagated by the matrix operations shows that if $\|F_x\|$ is reasonably well conditioned, then these errors may be reduced further by reducing the spacing, i.e., η. However, there is no guarantee that these errors will not increase. Consequently, the irreducible error in the gradient $\nabla_f G$ will prevent accurate determination of P_t, although no difficulty should arise in the determination of $P_{t,\alpha}$.

Determination of an Initial Design

The gradient methods may be applied to the solution of the system design problem once an initial design (i.e., the point P_o) has been determined. Setting the variables $\mathbf{z}'$ equal to zero according to the previous discussion immediately fixes the values of the variables $\mathbf{v}$ which are subject to type 4 and 5 constraints. The remaining variables may then be chosen arbitrarily. Let the point so determined be P_o', which can be taken as the point P_o only if

$$\phi_0' \leqslant c_f \tag{40}$$

If equation 40 is not satisfied, then the error in the constraints must be reduced.

This may first be attempted by employing the Newton-Raphson iteration (equation 13) several times, until equation 40 is satisfied. If the iteration proves to be divergent, so that the error $\mathbf{y}$ increases rather than decreases, then a least-squares minimization must be employed; i.e., the function

$$H_f(\mathbf{v}, \mathbf{z}) = \mathbf{y}, \mathbf{y} \tag{41}$$

must be minimized. The gradient of this unconstrained function is simply

$$(\nabla H_f)_t = 2(\mathbf{y})_t \| F_x : F_w \| \tag{42}$$

so that $H_f(\mathbf{v},\mathbf{z})$ may be minimized by the techniques described herein (note that the calculations involving ϕ have no significance for an unconstrained problem). Notice that if the minimum value of $H_f(\mathbf{v},\mathbf{z})$ is sufficiently greater than zero such that equation 40 is not satisfied, then a solution to the optimization problem cannot be obtained for $\mathbf{z}'=0$ by any method without either changing c_f or some of the bounds (α or β).

Conclusions

The gradient design technique for solving system optimization problems consists primarily of a group of criteria for governing the application of gradient minimization procedures. These criteria have been selected so as to guarantee convergence to a "solution" (i.e., a relative minimum) and, as far as possible, to achieve that solution rapidly. Where necessary, they have been expressed in terms significant to the designer. A simplified version of this design technique has been coded for a digital computer[1] and the results obtained strongly indicate the need for the complete set of the procedures described herein.

These gradient methods may be applied to any problem which meets the following restrictions:

1. The criterion function $G(\mathbf{v})$ and the constraint functions $g_i(\mathbf{v})$ must be continuous functions of continuous variables.

2. The number of constraints must be less than the number of variables ($m<n$).

A valid solution can be obtained for any problem which meets the following, additional restrictions:

3. The Jacobian at a relative minimum must be nonzero ($J^t \neq 0$).

4. The second-order behavior of the explicit expression for the criterion function must dominate the higher-order behavior near a relative minimum ($\|L^*\|^t \simeq$ constant $\neq 0$).

5. The error in the evaluation of the gradient must be small.

The problem of designing an optimum system may be solved by the gradient methods if the problem meets both the foregoing and the following restrictions:

6. The optimization problem must be expressible as a problem in minimization of a function subject to constraints.

7. The system topology must remain fixed during the minimization.

(Note that restriction 7 is essentially the same as restriction 1.) The majority of system design problems can be phrased so as to meet restrictions 1, 2, 6, and 7, and in most of these cases, restrictions 3, 4, and 5 will also be met. Thus, the gradient methods may usually be employed successfully for the design of optimum systems. For further details about these methods, including a computer block diagram, see reference 6.

References

1. A Generalized Computer Procedure for the Design of Optimum Systems—Part I, **R. Roderick Brown.** *AIEE Transactions*, pp. 285–89 of this issue.

2. System Synthesis With the Aid of Digital Computers, **J. B. Dennis, R. F. Nease, R. M. Saunders.** *Ibid.*, vol. 78, pt. I (*Communication and Electronics*), 1959, *Paper no. 59-196*).

3. Gradient Methods of Maximization, **J. B. Crockett, H. Chernoff.** *Pacific Journal of Mathematics*, Tokyo, Japan, vol. 5, 1955, pp. 33–50.

4. Acceleration of the Optimum Gradient Method, Preliminary Report (abstract), **G. E. Forsythe, T. S. Motzkin.** *Bulletin*, American Mathematical Society, Providence, R. I., vol. 57, 1951, pp. 304–05.

5. Nonlinear Programming, **H. W. Kuhn, A. W. Tucker.** *Proceedings*, Second Berkeley Symposium on Mathematical Statistics and Probability, University of California, Berkeley, Calif., 1951, pp. 481–92.

6. Gradient Methods for the Computer Solution of System Optimization Problems, **R. R. Brown.** *WADC Technical Note 57-159*, Wright Air Development Center, Dayton, Ohio, Sept. 1957.

The Design of the Corner-Reflector Antenna

H. P. NEFF, JR.
NONMEMBER AIEE

J. D. TILLMAN
ASSOCIATE MEMBER AIEE

THE CORNER-REFLECTOR antenna has been investigated by Kraus and others.[1,2] The mathematical expressions for the array factor as given in this earlier work are cumbersome, making it difficult to design such an antenna accurately. It is the purpose of this paper to overcome this difficulty.

The method of images may be used to analyze the corner-reflector antenna for corner angles $\alpha=\pi/n$, where n is any integer other than zero. For any such corner angle the antenna and its images in the reflecting sheets form a virtual circular array of $2n$ elements with π radians of current phase shift between adjacent elements in the array. When the array factor is derived as a special case of the general circular array, much simpler mathematical expressions are obtained, making the design problem a simple one. The results of Kraus will reduce to those derived in this paper.

Derivation of the Expression for the Array Factor

A corner-reflector antenna consists of two conducting planes which intersect along some apex, and some driven radiating element, usually a dipole, located equidistant from the perfectly conducting planes. The only situation which is amenable to mathematical analysis is that in which the conducting planes are infinite in extent, and in which the angle between the planes is π/n, with n being any positive integer. Fig. 1 shows the case of $n=3$. For the case where $n=3$, the driven element and its images in the conducting sheets form a virtual circular array with $2n$ elements, and a uniform progressive phase shift between the elements of π radians.

The analysis of such a circular array is

Paper 59-87, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 14, 1958; made available for printing December 12, 1958.

H. P. Neff, Jr. and J. D. Tillman are with the University of Tennessee, Knoxville, Tenn.

The research reported in this paper has been supported by the Air Force Cambridge Research Center under Contract no. *AF 19(604)-1557*.

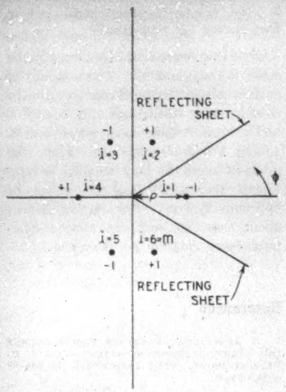


Fig. 1. The equivalence of the corner-reflector antenna array and the circular array in the plane $\theta=\pi/2$ for the case n=3

known.[3] Consider m parallel dipoles, equally spaced around the circumference of a circle of radius ρ. Let the center of the circle be located at the center of a system of spherical co-ordinates, as shown in Fig. 2. If the current in the ith element is

$$I_i=\frac{I_0}{m}\epsilon^{j\left[\omega t+\left(\frac{2\pi}{m}\right)ni\right]} \tag{1}$$

there is a uniform progressive phase shift between the elements of $2\pi/mn$ radians, and total phase shift around the ring of $2\pi n$ radians. The field from the ith element is

$$E_i=\frac{I_0}{m}\epsilon^{j\left[\omega t+\left(\frac{2\pi n}{m}\right)i+\frac{2\pi\rho}{\lambda}\sin\theta\cos\left(\phi-\frac{2\pi}{m}\right)i\right]} \tag{2}$$

The array is then found by summing E_i from 1 to m. This summation is carried out in reference 2. The result is

$$E=j^nI_0J_n(z)\epsilon^{j(\omega t+n\phi)}+I_0\sum_{k=1}^{\infty}\times$$
$$\{j^{km-n}J_{km-n}(z)\epsilon^{j}[\omega t-(km-n)\phi]+ j^{km+n}J_{km+n}(z)\epsilon^{j}[\omega t+(km+n)\phi]\} \tag{3}$$

where

$$z=\frac{2\pi\rho}{\lambda}\sin\theta.$$

For the case of the corner reflector, the number of elements in the virtual array is $2n$, so we specify that $m=2n$, which requires that m is even. The phase difference between adjacent elements is $2\pi n/m=2\pi n/2n=\pi$, which is correct for the virtual array. The array factor can now be simplified to

$$E=2I_0\epsilon^{j\omega t}\left\{j^nJ_n(z)\cos n\phi+\sum_{k=1}^{\infty}j^{n(2k+1)}J_{n(2k+1)}(z)\cos(2k+1)n\phi\right\} \tag{4}$$

as is shown in the Appendix. This expression gives the radiation pattern of the corner reflector, subject of course, to the restriction that the corner angle is π/n, and that the conducting sheets are infinite in extent.

As an example of the method, let it be required to design a corner-reflector antenna with a corner angle of 60° ($n=3$). The array factor becomes

$$E=2I_0\epsilon^{j\omega t}\{j^3J_3(z)\cos 3\phi+j^9J_9(z)\cos 9\phi+ j^{15}J_{15}(z)\cos 15\phi+\ldots\ldots\} \tag{5}$$

In the plane of the array

$\theta=\pi/2$, and

$$z=\frac{2\pi\rho}{\lambda}.$$

For a one-lobed pattern then the designer must choose a radius, ρ, or antenna-to-apex spacing such that $J_9(z)$, $J_{15}(z)$, etc.

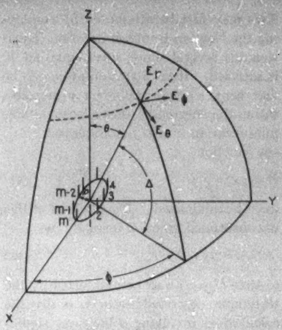


Fig. 2. Co-ordinate system used for analysis of ring arrays

are negligible compared to $J_3(z)$. This condition is usually satisfied if $z\leqq n$ due to the rapid decrease in magnitude of the higher order Bessel functions. For the example above

$$\frac{2\pi\rho}{\lambda}=3$$

$$\frac{\rho}{\lambda}=\frac{3}{2\pi}=0.477$$

$$J_3(3)=0.30906$$

$$J_9(3)=0.00008$$

$$J_{15}(3)=2.9\times10^{-9}$$

All terms except the first in equation 5 may be neglected leaving $E=k\cos 3\phi$ where k is a constant.

In order to obtain an array factor with a maximum magnitude in the direction $\phi=0$, $\theta=\pi/2$, it is necessary to require that

$$\rho\leqq 0.699\lambda \text{ or } z\leqq 4.2\ (n=3),$$

since $J_3(z)$ is maximum for this value of z. As the value of ρ is decreased below

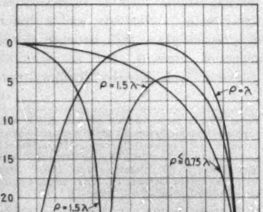
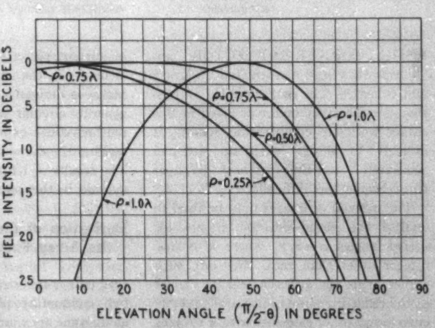


Fig. 3 (left). Array factor for corner angle of 60°, ($n=3$), in the plane $\theta=\pi/2$

Fig. 4 (right). Array factor for corner angle of 60°, ($n=3$), in the plane $\phi=0$



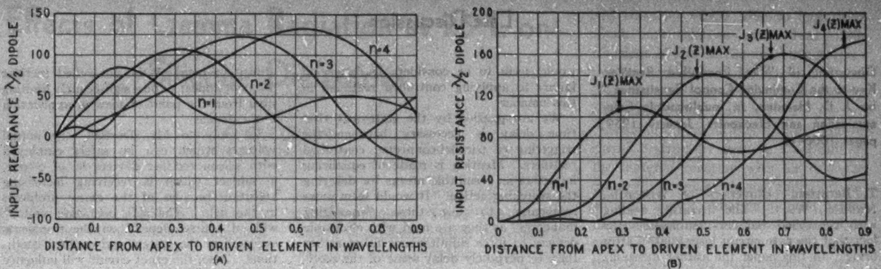


Fig. 5. **Input resistance and reactance in ohms with λ/2 dipole as the driven element**

A—Resistance B—Reactance

the value given in the preceding statement, the shape of the array factor remains substantially the same. The magnitude of the array factor decreases approximately as $J_3(z)$ with z decreasing. If the driven element of the corner reflector is a half-wave dipole at $\rho=0.669\lambda$, then its input resistance will decrease as ρ decreases. With constant input power to the antenna terminals it should be possible to maintain approximately constant field strength in the direction $\phi=0$, $\theta=\pi/2$. In this manner the corner-reflector antenna may be a broad-band antenna. The array factor as a function of antenna-to-apex spacing is shown in Fig. 3 and Fig. 4 for the cases $\theta=\pi/2$ and $\phi=0$, respectively.

Thus, for any corner-reflector antenna with corner angle π/n the array factor becomes

$$E=KJ_n\left(\frac{2\pi\rho}{\lambda}\sin\theta\right)\cos n\phi+\text{residual terms}$$

The designer must merely choose a radius such that the residual terms may be neglected. This is the expression for the pattern of an array of isotropic radiators, and must be multiplied by the expression for the pattern of the driven element to obtain the equation of the final pattern. Obviously, no radiated field will exist outside the corner angle under the original assumptions.

The input impedance of the driven element may be calculated very simply due to the symmetry of the array. In general, as shown by Kraus,[1,2]

$$Z_{in}=Z_{11}+\sum_{i=1}^{n}\epsilon_i\cos\pi iZ_{1\,i+1} \qquad (6)$$

where $\epsilon_i=1$ when $i=n$, and $\epsilon_i=2$ otherwise. For the case above ($n=3$).

$$Z_{in}=Z_{11}-2Z_{12}+2Z_{13}-Z_{14}$$

Results are shown in Fig. 5 for a half-wave dipole as the driven element.

Conclusions

The results given here for the corner-reflector antenna relate the pattern to the spacing from the apex to the driven element in a much clearer manner than the previous analyses. A satisfactory design can be obtained by choosing this spacing so as to minimize the residual terms. Maximum field for a given size current occurs if this spacing is chosen to maximize the factor $Jn(2\pi\rho/\lambda\sin\theta)$.

Appendix

Derivation of Expression for Array Factor

Equation 3 gives the array factor of a circular array of m element, with a progressive phase difference of $2\pi/mn$ between adjacent elements. For the virtual array formed by a corner reflector $m=2n$, and equation 3 becomes

$$E=j^nI_oJ_n(z)\epsilon^{j(\omega t+n\phi)}+ I_o\sum_{k=1}^{\infty}\left\{j^{(2k-1)n}J_{(2k-1)}(z)\epsilon^{j[\omega t-(2k-1)n\phi]}+ j^{(2k+1)n}J_{(2k+1)n}(z)\epsilon^{j[\omega t+(2k+1)n\phi]}\right\} \qquad (7)$$

The first term of the summation is $I_oj^nJ_n(z)\epsilon^{j(\omega t-n\phi)}$, and this can be combined with the leading term to give

$$2j^nI_oJ_n(z)\cos n\phi\epsilon^{j\omega t}$$

In the same way the second term of the summation is

$$j^{3n}I_oJ_{3n}(z)\epsilon^{j(\omega t+3n\phi)}$$

resulting from $k=1$, and the third term, coming from $k=2$ is

$$j^{3n}I_oJ_3{}^n(z)\epsilon^{j(\omega t-3n\phi)}$$

These combine to give

$$j^{3n}I_oJ_{3n}(z)\cos 3n\phi\epsilon^{j\omega t}$$

Then equation 7 can be written

$$E=2I_o\epsilon^{j\omega t}\left[j^nJ_n(z)\cos n\phi+ \sum_{k=1}^{\infty}\epsilon j^{(2k+1)}J_{(2k+1)}(z)\cos(2k+1)n\phi\right] \qquad (8)$$

This is equation 4 of the text.

References

1. ANTENNAS (book), J. D. Kraus. McGraw-Hill Book Company, Inc., New York, N. Y., 1950, pp. 328–36.

2. THE CORNER REFLECTOR ANTENNA, J. D. Kraus. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 28, Nov. 1940, pp. 513–19.

3. THE USE OF A RING ARRAY AS A SKIP RANGE ANTENNA, J. D. Tillman, W. T. Patton, C. E. Blakely, F. V. Schultz. *Ibid.*, vol. 43, Nov. 1955, pp. 1655–60.

Late Discussion

Discussion of "The Controlled Rectifier: Key to the Continuing Control Renaissance," by J. D. Harnden, Jr., published in Communication and Electronics, January 1959, pages 1006–12.

T. Bernstein (University of Wisconsin, Madison, Wis.): The literature dealing with controlled rectifiers emphasizes the fact that the gate circuit loses control after firing. In this paper the author mentions that there is a possibility that the controlled rectifier can be turned off by gate action. Have such units been made as yet? What would be the magnitude of the gate signal necessary to turn off the rectifier as compared to the signal necessary to turn it on? Since the controlled rectifiers presently available are not designed for reverse voltage on the gate junction, I am in doubt as to whether the turn-off gate voltage would be a reversed polarity gate signal. This information will be useful as an indication of the future potentialities of the controlled rectifier.

I. K. Dortort (I-T-E Circuit Breaker Company, Philadelphia, Pa.): In arriving at some estimate of the efficiency of equipments using the controlled rectifier *ZJ-39A*, the author points out that the forward drop is but one volt out of a peak voltage rating of 400. We assume the peak voltage rating is equivalent to the PIV (peak inverse voltage) rating of power diodes, in which case it is expected that the recommended rms alternating voltage to be applied to these rectifiers would not be much over 115 volts. Even at this reduced voltage, the efficiency is still excellent. However, the author's statement raises the question of recommended peak operating voltage. This question is further emphasized by the statement that it is possible to use these devices in a control system in which forward conduction is initiated by exceeding the breakover voltage instead of a signal on the control electrode. Since the forward breakover voltage and the PIV rating of the rectifier are stated to be approximately equal, there would seem to be considerable danger of failure in applying controlled rectifiers in this manner.

We are puzzled by the statement that time delay is necessary between the triggering of parallel-connected controlled rectifiers. Mention is made of equalizing inductors as a suitable means for delaying the switching action. It would seem more reasonable to use these "equalizing inductors" as they are used with electronic devices to force simultaneous firing rather than to purposely delay some of the rectifiers and thereby cause them to carry less than their proportionate share of load.

The bridge-connected rectifier of Fig. 2(H) is claimed to have inherent commutation protection. I do not understand what is mean there: Would the author explain?

J. D. Harnden, Jr.: In reply to Mr. Bernstein, at present, the controlled rectifier is primarily intended for operation in circuits in which turn-off is accomplished by commutation. This implies that the device is not intended to be operated in a turn-off direction by gate action. With the present production *C35* device it is possible to turn off anode-cathode current flow by diverting a portion of the current out through the gate circuit. However, this action is only possible in the relatively low-current region such as 500 milliamperes or less. Under such operation, the current flow in the gate lead is considerably greater than required to initiate conduction through the controlled rectifier. In evaluating the merits of a device which can both turn on and turn off a circuit, it becomes necessary to consider not only resistive but reactive loads.

It is, of course, necessary to consider the transient energy involved in opening a reactive circuit. Some provision must be made to protect the semiconducting device, in addition to absorbing its energy. Generally speaking, this involves the use of additional energy absorbing or protecting equipment. In most instances, it appears that this requirement is no less stringent than that involved in the commutation of a thyratron-like device. Therefore, the ability to turn off the controlled rectifier does not appear to be a significantly limiting item from the point of view of application of the device.

In reply to Mr. Dortort, the actual voltages which can be safely employed with silicon rectifier devices is, of course, a subject which is receiving increasing attention because of the many problems in the field. The safe operating voltage will, of course, depend on the transients expected in the system and on load conditions. Also, the exact circuit will influence the safe voltage for operating the devices.

The statement with regard to turning on the controlled rectifier by breakover voltage instead of a signal on the control electrode or gate implies, of course, that adequate protection methods will be used, so that the peak inverse rating of the device is not exceeded. This may take the form of an auxiliary series rectifier so as to extend the PIV rating. However, in general, this method of application will be used with unidirectional pulses so there is no danger with respect to the reverse voltage.

The statement concerning turn-on of parallel-connected devices is based primarily on the fact that there is a finite and measureable turn-on of the controlled rectifiers in addition to a measurable and finite delay time. Therefore, since these times are widely different in controlled rectifiers, it may be possible for one unit to be turned on sufficiently to prevent the other controlled rectifier from turning on. This means that the first unit to be completely turned on may act as a clamp or switch with regard to the other units in the system. The reference does not imply that there is any difference in the application of the gate pulse to all of the units which may be operating in parallel. Generally, however, it appears that the leads normally employed in connecting the controlled rectifiers possess sufficient inductance so that the turning-on action is almost inherently equalized.

The reference to commutation protection of the bridge rectifier circuit refers to operation with inductive loads, etc., wherein the normal diodes provide a free-wheeling path, thus protecting the controlled rectifiers from transient disturbances.

Effects of Gamma Radiation at 25 C on Silicone Dielectrics

C. G. CURRIN
ASSOCIATE MEMBER AIEE

ELECTRIC EQUIPMENT such as pump motors, instrumentation, control apparatus, and wire and cable, is necessary for the operation of nuclear reactors in central stations and in marine, rail, and aeronautical transportation equipment. For engineering and economic reasons, some of the electric apparatus is placed close to the reactor in regions of strong nuclear radiation. So that equipment can be designed to be reliable when exposed to this radiation, engineers require data on the behavior of all the components and materials when subjected to this environment.

To obtain radiation-effects data on one group of dielectrics, measurements of effects of gamma radiation on physical and dielectric properties of liquid, semisolid, elastomeric, and resinous silicones are being made. These measurements indicate that many silicone dielectrics have appreciable resistance to radiation exposure.

Basic Radiation Effects

The radiation exposure to which electric apparatus is subjected varies with the energy rating of the reactor, the shielding between the reactor and the apparatus, the geometry of the reactor and equipment, and the expected life of the apparatus in terms of reactor operating time. Such exposures may vary from 10 to 10,000 megarads.[1]

Nuclear radiation affects some types of materials much more than other types. Certain solid-state devices, such as transistors and rectifiers, are damaged severely by relatively small radiation exposures. Other crystalline materials such as iron, copper, and mica are many times more resistant. Organic or polymeric materials, used primarily as electrical insulation in electrical equipment, are intermediate in resistance to radiation. The radiation exposure necessary to damage insulating materials varies considerably since it is dependent upon the composition, type, and application of the insulation.

The primary mechanisms by which radiation affects polymeric materials have been shown to involve ionization and excitation.[2,3] This is true for each of the major types of radiation, beta, gamma, and neutron, found near nuclear reactors. The degree of ionization or excitation is proportional to the total radiation energy absorbed by the material. The fundamental unit for absorbed radiation energy is the rad (1.0 rad = 100 ergs absorbed/gram material). A more convenient unit for describing effects on dielectric materials is megarad (rads$\times 10^{-6}$). The AIEE Subcommittee on Effects of Radiation on Insulation has adopted the rad and megarad as standard units of irradiation dose for reporting radiation effects. Radiation doses required to affect polymeric materials are usually from 1.0 to 1,000 megarads.

Ionization and/or excitation results when chemical bonds in polymeric materials are broken by absorbed radiation energy. The resulting ionic species usually recombine forming new and often different chemical bonds. The structure of the polymer is thereby altered, changing the physical properties of the material. Depending upon the manner in which the chemical bonds are broken and reformed, the material becomes either brittle or soft.

Changes in the chemical structure of a polymer also modify the dielectric properties. Furthermore, minute traces of radiation-induced impurities have a sizeable effect on these properties. Some of these impurities are ionized atoms, others are electrons freed by the radiation from the parent atoms. Still other impurities are ionized fragments of the original polymer or new stable compounds.

Excluding the stable compounds, these impurities have a limited life before combining with one another or with the material. Therefore the effects of radiation on dielectric properties are of two types: permanent effects, due to the stable impurity compounds and rearrangement of the chemical structure, and transient effects, due to impurities which have relatively short lives. The transient effects on the dielectric properties are dominant during irradiation and also, for some materials, immediately after irradiation.[4]

Some of the impurities or alterations in chemical structure may be the result of the material reacting with air or some other environment during radiation exposure. Hence radiation-induced oxidation or other similar chemical reactions may be the cause of some of the radiation effects. This environmental factor is analogous to effects of high temperature on the stability of polymeric materials.

The temperature of the polymeric material during radiation exposure influences the mobility of the ionized impurities and thus affects the rate at which radiation effects are produced. The intensity of radiation has also been shown to be a factor in at least one material, as discussed in an unpublished paper by Jerome Goodman and J. H. Coleman.

The amount of high-energy beta or gamma radiation absorbed by a polymeric material, which determines the degree of ionization, is primarily a function of the density. In a neutron field, however, the energy absorbed is largely determined by the atomic composition of the material. Polymeric materials are usually composed of hydrogen, carbon, oxygen, and in some cases nitrogen and silicon atoms. All of these atoms have relatively small neutron absorption coefficients, yet those polymeric materials containing boron, lithium, or similar additives with high neutron absorption coefficients, may be expected to absorb much more energy from a neutron field.[5]

Silicone Insulation Evaluation Program

A continuing program was initiated to measure the major effects of radiation on the physical and dielectric properties of a number of typical silicone insulating materials. Silicones are noted for reliability and performance in severe temperature and moisture environments; this program is to evaluate these dielectrics under a third environmental condition, nuclear radiation.

This test program involves only silicone materials and simple combinations of these dielectrics with certain inorganic materials. Since the immediate environment of a material may be a major factor in determining the radiation resistance of a material, a separate program is necessary to evaluate the resistance

---

Paper 59-117, recommended by the AIEE Electrical Insulation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 10, 1958; made available for printing January 13, 1959.

C. G. Currin is with the Dow Corning Corporation, Midland, Mich.

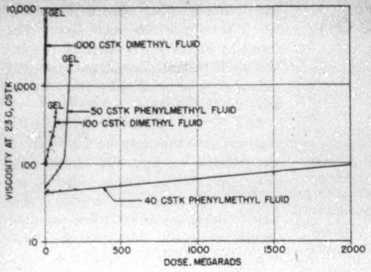
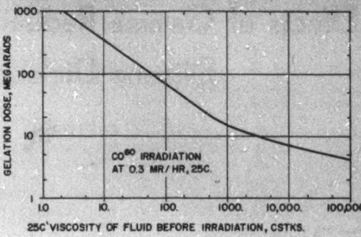


Fig. 1 (left). Viscosity of irradiated silicone fluids

Fig. 2 (right). Radiation dose required to gel linear dimethyl silicone fluids



of silicone insulation systems. Four types of silicone dielectrics are studied: liquids, compounds, elastomers, and resins.

Samples of these materials are exposed to cobalt 60 gamma radiation. Cobalt 60 emits rays of 1.17 and 1.33 million electron volt (Mev) energies. The environment around the multikilocurie hollow cylindrical source used is air at atmospheric pressure and room temperature.

Samples of silicone liquids are placed in 4-ounce glass bottles; the bottles are vented by a small hole in each cap to allow the escape of gases evolved from the liquids during irradiation. The samples of compounds are placed in 2-ounce tin ointment cans; samples of elastomers and solids are either without a protective container or are wrapped in 0.006-inch thick aluminum foil. However, no significant evidence has been found that any of these container materials affects the radiation damage mechanism.

To determine the irradiation dose received by the samples, calculations are made on the basis that the samples absorb the same energy from the radiation field as water would which is the normal technique for polymeric materials.[6]

The dose rate, or rate of energy absorption, is determined by the strength and configuration of the source and the distances between the source and the sample. Dose rates from 0.02 to 0.7 megarads/hour (per hour) are obtained based on ferrous sulfate dosimetry. However, for most of the exposures, the dose rate is between 0.3 and 0.7 megarads/hour.

The spatial distribution of the radiation intensity and the size and orientation of the sample largely determine the consistency of the irradiation dose over the volume of the sample. For this reason, only thin samples are placed in regions of substantial radiation-intensity gradient. The maximum difference in dose between two regions of any sample does not exceed 25%.

Basic physical and dielectric properties of the samples are measured after the radiation exposures. An exponential decay of certain properties with time after exposure have been reported for some materials.[4] The time constants of the decays observed with silicone insulation have been less than 10 minutes; thus the properties are essentially stable one hour after irradiation. Even so, measurements are usually not made until at least 24 hours after the radiation exposure.

The physical properties measured vary with the physical form of the materials. In general, however, obvious changes in appearance are recorded and a measurement of the hardness of a solid or viscosity of a liquid is made. Hardness or viscosity is one indication of alteration in chemical structure by the irradiation.

Four major dielectric properties of the irradiated samples are measured. Electric strength, volume resistivity, dielectric constant, and dissipation factor are measured at 23 C (degrees centigrade) or 25 C. Arc resistance of some of the irradiated samples is also measured.

The four major dielectric properties of several representative materials are also measured during exposure to gamma radiation. Dose rates from 0.02 to 0.5 megarads/hour are employed in these tests. Special techniques are used to make the resistivity and complex permittivity measurements, since air surrounding the sample is highly ionized by the radiation.

It was found necessary to submerge the samples and the guarded electrode system, including the cable connections, in a dimethyl silicone liquid dielectric to minimize effects of the ionized air. Although the silicone liquid is also ionized, the dielectric losses are much less than those of the ionized air.

Descriptions of the test methods, equipment, and samples are detailed in the Appendix.

The data being obtained during the evaluation program are statistically analyzed when practical to obtain the best estimate of the radiation effects and the statistical confidence limits thereof. Confidence limits at the 95% level are used; that is, there was no more than a 5% probability that each significant conclusion was made in error.

Radiation Effects on Liquids

The effects of gamma radiation on silicone fluids are demonstrated by data on four different Dow Corning Corporation fluids:

Silicone Fluids	Centistokes(Cs)
A—Dimethyl, 200	100
B—Dimethyl, 200	1,000
C—Phenylmethyl, 510	50
D—Phenylmethyl, 704	40

The first two are typical silicone fluids used as liquid dielectrics in capacitors and liquid immersed equipment. They differ primarily in viscosities.[7] The 50-cs phenylmethyl silicone fluid contains only a small number of phenyl groups. The 40-cs phenylmethyl silicone fluid is a low-molecular-weight polysiloxane with approximately equal numbers of phenyl and methyl groups.

PHYSICAL PROPERTIES

Fig. 1 shows that exposure to gamma radiation increases the viscosities of all these liquids indicating an increase in molecular weights. The figure shows that the increase, in general, follows the equation

$$\log\left(\frac{\nu}{\nu_0}\right)=\beta_\nu D \qquad (1)$$

when ν is the viscosity after irradiation, ν_0 is the viscosity prior to radiation exposure, β_ν is the coefficient of viscosity increase, and D is the irradiation dose. As the dose absorbed by silicone fluids increases beyond a critical value, the rate of increase of viscosity is sometimes greater than that indicated by equation 1. After a dose approximately 50% greater than this critical dose, the fluid gels and becomes a soft rubbery solid.

Fig. 1 also shows that the rate of viscosity increase is markedly different among these four liquids. This rate of viscosity increase, β, in equation 1, is tabulated for the four fluids.

Silicone Fluid, Cs	Approximate Phenyl/Methyl Ratio	β_v, Megarad^{-1}
100, dimethyl	0.00	0.0079 ±0.0013
1,000, dimethyl	0.00	0.063 ±0.032
50, phenylmethyl	<0.1	0.0026 ±0.0001
40, phenylmethyl	1.00	0.00017 ±0.00002

Dimethyl silicone fluids with high viscosities are the least resistant to radiation effects on viscosity. Also, the radiation resistance increases with an increasing proportion of phenyl to methyl groups in the fluid.

After a dose usually greater than that required to double the viscosity, the fluid will gel. The irradiation dose required to gel dimethyl silicone fluids has been found in the following manner.

It has been shown that the dose required to gel a liquid is

$$D_{gel}=\frac{481.6\times10^{3}}{G\ Mw}\ \text{megarads} \qquad (2)$$

when G is the number of chemical bonds created by 100 electron volts of irradiation energy and Mw is the average molecular weight.[8]

Experiments were made to determine the dose required to gel linear dimethyl silicone fluids of 100- and 1,000-cs viscosity. It was found that at a dose rate of 0.3 megarad/hour and exposure at 25 C, on an average 65±5 megarads would gel 100-cs fluid and 14.0±2.0 would gel 1,000-cs fluid. Obtaining the molecular weights of these fluids by two separate methods, equation 2 can be used to estimate the value of G for linear dimethyl silicone fluids. The calculations are tabulated in Table I.

Substituting the best estimate of G for these linear dimethyl silicone fluids, and assuming that G does not vary appreciably with molecular weight, equation 2 may be reduced to

$$D_{gel}=\frac{423\times10^{3}}{Mw}\ \text{megarads} \qquad (3)$$

Calculated gelation doses for dimethyl silicone fluids having room temperature viscosities from 2.0 cs to 100,000 cs are plotted in Fig. 2. The gelation doses for fluids of this viscosity range vary from 1,100 megarads for the 2.0-cs fluid to 4.0 megarads for the 100,000-cs fluid.

Table I. Calculations for Value of G

Dimethyl Silicone Fluid Viscosity, Cs	D_{gel}, Megarads	Mw, Avg.	G, Bonds Formed per 100 Electron Volts
1,000	14	32,500*	0.98
	14	26,400†	1.30
100	65	6,500*	1.14
	65	6,600†	1.12
		Average	1.14±0.21

* For authority for *Mw* quantities, see ref. 9.

† For authority for *Mw* quantities, see ref. 10.

This value of G is for gelling only dimethyl silicone fluids. Sufficient data on the gelation of phenylmethyl silicone fluids are not available; hence, the G value for most of these liquids can not be calculated. However, since the presence of phenyl groups greatly increases the resistance of silicone fluids to viscosity changes and gelation, the G value of these fluids is expected to be much less.

Other physical properties of silicone fluids are also affected by radiation exposure. The densities increase slightly and are proportional to the increases in viscosity. However, these increases are slight and are of little engineering significance. In the case of the 50-cs phenylmethyl silicone fluid, the density increases by 1.0% after 140 megarads of irradiation, the dose required to double the viscosity. There is also a similar small increase in the refractive index.

The color of the dimethyl silicone fluids remains water-white during irradiation, even after gelation takes place. The color of the phenylmethyl silicone fluids becomes amber upon irradiation and darkens after heating the fluid to 250 C for 4 hours.

Of major engineering importance, however, are the gases evolved from silicone liquids during irradiation. Although no measurements of the quantity or composition of the gases which evolve have been made in this study, previously published data indicate that three different gases are produced:

Component Gas	Quantity, Volume Per Cent
Hydrogen	34.4
Methane	61.0
Ethane	4.6

These are mass-spectrometer data of the gases evolved from a dimethyl silicone fluid irradiated 50 megarads by a cobalt 60 gamma radiation source.[11]

Considering the chemical structure of the silicone fluids and of these simple gases, probably the same mechanism which increases the viscosity also produces these gases; this mechanism is known as cross-linking.

Considering the cross-linking mechanisms probable, the number of molecules of gas produced, regardless of the composition, is approximately equal to the number of cross-links formed in the polymer by irradiation. However, this approximation may be in error by a factor or two or slightly more.

This approximation can be used to obtain an estimate of the quantity of gas evolved. The number of cross-links formed by the gelation dose can be derived from the G value previously calculated. Remembering that this is an average value of G over the total dose required to gel the fluid, it can be shown that 1 megarad of irradiation will produce $6.3\times10^{17}\,G$ crosslinks in a silicone liquid.

Using the gelation value of G for dimethyl fluids, 1.14, calculations show that 0.03 milliliter of gas at 0 C and 760 millimeters pressure are evolved from one gram of fluid by a dose of one megarad. Considering the range of approximation made, the quantity of gas evolved by 1 megarad can be estimated to be between 0.01 and 0.05 milliliter per gram of fluid. Since the value of G does not vary appreciably with the different viscosity grades of these fluids, all dimethyl fluids can be expected to have the same gas evolution rate.

This estimate of gas evolution indicates that an exposure of 50 megarads will produce gas with a volume approximately 150% of the parent dimethyl fluid. The phenylmethyl silicones, which have much smaller values of G, produce substantially less gas.

In sealed equipment, such as silicone impregnated capacitors, this gassing may be the major detrimental effect of radiation exposure. To insure reliability, such components or equipment must be provided with a safe method for venting the gas.

DIELECTRIC PROPERTIES

There are two basic types of effects of irradiation on the dielectric properties: the permanent effects observed after irradiation and the transient effects found during and sometimes immediately after the radiation exposure. These two types of effects were studied separately.

Permanent Effects. Irradiation causes a small increase in the dielectric constants of silicone fluids. This increase at frequencies of 100 cps (cycles per second) and 0.1 mc/sec (megacycles per second) is shown in Fig. 3(A) for two types of fluids. The effects on the 1,000-cs dimethyl fluid are similar to those shown for the 100-cs dimethyl fluid. The increase indicates the formation of polar groups, since cross-linking of silicone fluids has no appreciable effect on the dielectric constant.

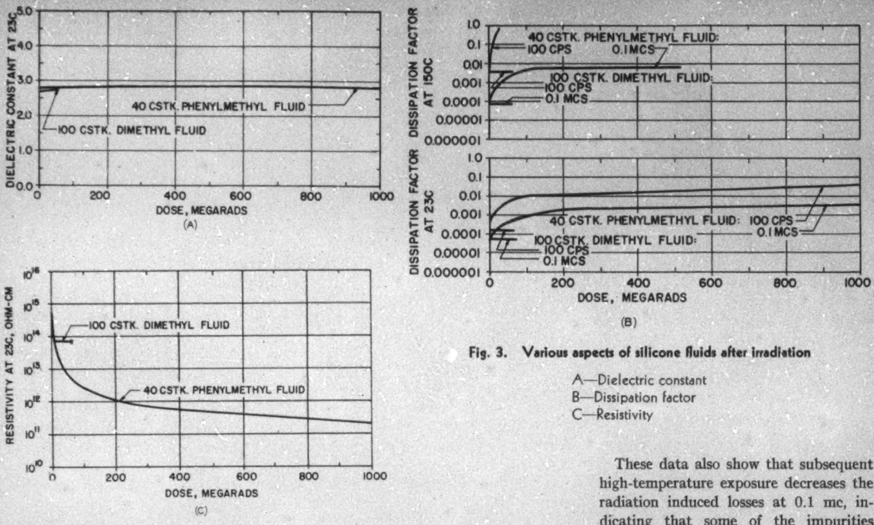


Fig. 3. Various aspects of silicone fluids after irradiation

A—Dielectric constant
B—Dissipation factor
C—Resistivity

The increase in the room temperature dielectric constants of silicone fluids is proportional to the dose. Thus, the dielectric constant can be mathematically related to the dose by the equation

$$K = K_0 + \beta_k D \tag{4}$$

when K is the dielectric constant of the irradiated fluid, K_0 is the dielectric constant prior to irradiation, β_k is the coefficient of increase in dielectric constant, and D is the dose. Values of β_k for the 25 C 0.1-mc dielectric constant for three silicone fluids are tabulated in the following:

Silicone Fluid, Cs	β_k, Megarads^{-1}
100, dimethyl	0.006 ±0.0004
50, phenylmethyl	0.00043±0.00008
40, phenylmethyl	0.00012±0.00003

Again the phenylmethyl silicone fluids are more resistant to radiation effects.

As expected, the dielectric constants at 150 C are lower, approximately 2.4, and are less affected by radiation exposure.

There is a major difference between irradiated dimethyl and phenylmethyl silicone fluids, however, when the dissipation factor is considered. Fig. 3(B) shows radiation exposure has no appreciable effect on the dissipation factor of a dimethyl silicone fluid at frequencies from from 50 cps to 0.1 mc at either 23 C or 150 C.

However, as further shown in this figure, radiation exposure has a drastic effect on phenylmethyl silicone fluids. These fluids, which are otherwise the least affected by radiation, have relatively little resistance to effects on the dissipation factor.

The rate at which the dissipation factor of phenylmethyl fluids is affected decreases exponentially with increasing dose. This is unique. For almost all properties which have been measured on any material, the radiation effects are primarily a linear function of the dose. The exponential decrease indicates the mechanism by which the dissipation factor is affected is different from that which increases the viscosity; i.e., cross-linking. Since the rate decreases exponentially, this mechanism must be affected by the products of the process.

It is believed that the increase in dissipation factor is due to radiation damage to the phenyl group in phenylmethyl silicone fluids. This damage results in stable impurities, some of which are not volatile even at 225 C. This stability is indicated by the following data for the 40-cs phenylmethyl fluid after a dose of 100 megarads.

Test No.	Dissipation Factor at 23 C			
	Before Heating		After 4 Hours/250 C	
	100 Cps	0.1 Mc	100 Cps	0.1 Mc
1	0.0078	0.0026	0.033	0.00042
2	0.0079	0.0026	0.035	0.00042

These data also show that subsequent high-temperature exposure decreases the radiation induced losses at 0.1 mc, indicating that some of the impurities formed are either volatile or not thermally stable. The impurities formed by irradiation are not in sufficient quantity for detection by normal infrared analytical techniques.

Radiation exposure has a similar effect on the resistivity of silicone liquids, as is shown in Fig. 3(C) for one dimethyl silicone fluid and one phenylmethyl fluid. Radiation causes a small but statistically significant increase in the resistivity of dimethyl silicone fluid. Irradiation has a much greater effect on the phenylmethyl fluids; the rate of this is exponential. This indicates that the mechanism for this decrease in resistivity is probably the same as that causing the increase in dissipation factor.

The d-c electric strength of the fluid is damaged by exposure to gamma radiation. This damage occurs in both dimethyl and phenylmethyl fluids and is shown in Fig. 4 for the 40-cs

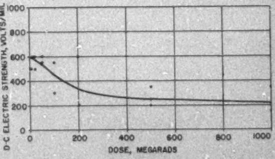


Fig. 4. Electric strength of 40-cs phenylmethyl fluid after irradiation; 0.020-inch gap between 0.50-inch-diameter spherical electrodes

Vol.
Pt. I
195

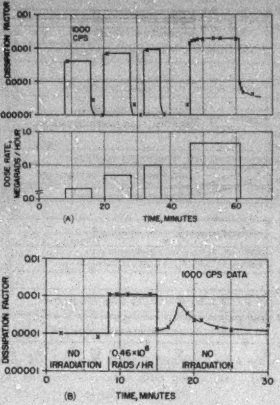


Fig. 5. Dissipation factor of silicone fluids during and after irradiation

A—100-cs dimethyl

B—50-cs phenylmethyl

phenylmethyl fluid. Irradiation reduces the average d-c electric strength to approximately one half of that of nonirradiated silicone fluids.

After irradiation the variations of the individual breakdown voltages increase. With the test equipment used, the coefficient of variation for nonirradiated fluids is less than 5%. The coefficient for the 40-cs phenylmethyl fluid data irradiated from 100 to 1,000 megarads is 45%; this difference is statistically significant at a high level. Thus, the practical electric strength of irradiated silicone fluids is considerably less than that indicated in Fig. 4. Results from the a-c measurements are similar to the d-c electric strength data.

Transient Effects. The intensity of radiation is the major factor in determining dissipation factor and resistivity during radiation exposure. Those effects which exist are transitory in nature and disappear immediately or shortly thereafter when the exposure is stopped.

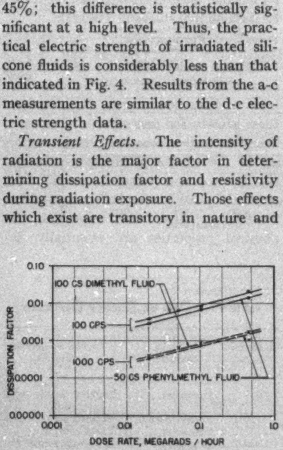


Fig. 6. Dissipation factor of two silicone fluids during irradiation

No significant transistory effect of irradiation was found on the dielectric constant of the 100-cs dimethyl fluid or the 50-cs phenylmethyl fluid.

The dissipation factor is grossly affected by the radiation exposure, however, as shown in Fig. 5(A) for the 100-cs dimethyl fluid and in Fig. 5(B) for the 50-cs phenylmethyl fluid. At dose rates greater than 0.1 megarad/hour, the 1,000-cps dissipation factor is a decade or two greater than prior to the irradiation. These figures also show an effect of time on the dissipation factor of both fluids although it is much more pronounced with the phenylmethyl fluid immediately after the irradiation ceases. In both fluids the time effect disappears within 15 minutes.

Measurements made after 16 hours of continuous irradiation did not indicate any time effects not evident during the first few minutes of irradiation.

At 100 cps, the dissipation factor during irradiation is even greater than at 1,000 cps as shown in Fig. 6. These data also indicate a linear relationship between the logarithm of dissipation factor and the logarithm of dose rate; the coefficient of this relationship is 0.50 for both types of silicone fluid measured at 100 and 1,000 cps.

A similar effect on resistivity is shown in Fig. 7 for the 100-cs dimethyl fluid. Data obtained on the 50-cs phenylmethyl fluid fall essentially on the same line. Here, however, the coefficient of the logarithm of resistivity-logarithm of dose-rate relationship is between 0.50 and 0.73.

Three measurements were made of the d-c electric strength of each type of silicone fluid during irradiation. The average breakdown voltage between 0.50-inch diameter spherical steel electrodes with a gap of 0.20 inch are tabulated in Table II.

During irradiation there is no evidence that the d-c electric strength is less than the strength when not exposed to radiation. The above data indicate the electric strengths of fluids previously exposed to radiation are greater during irradiation. The individul data also show that irradiation improves the consistency of the breakdown voltages of these previously irradiated fluids.

Radiation Effects on Compounds

Silicone compounds, manufactured by introducing substantial amounts of inorganic fillers into silicone fluids, are used as sealing compounds in electrical and electronic apparatus and components and as a water repellent surface coating for ceramic insulators.[12] The effects of gamma radiation on silicone compounds are demonstrated by data on two such compounds. Both compounds use silica fillers and differ primarily in the silicone fluids used.

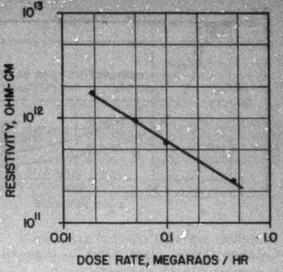


Fig. 7. Resistivity of 100-cs dimethyl silicone fluid during irradiation

Compound A—Dimethyl silicone fluid base, silica filler (Dow Corning 4 Compound)

Compound B—Phenylmethyl silicone fluid base, silica filler (Dow Corning 5 Compound)

Since inorganic fillers are relatively unaffected by radiation doses of 1,000 megarads or less, effects of radiation on these silicone compounds are due to radiation effects on the fluid base and possible radiation induced reactions between the fluid and filler. Many of the effects observed would be predictable knowing the basic characteristics of these silicone compounds and the previously discussed effects of radiation on silicone fluids.

PHYSICAL PROPERTIES

The consistency of these compounds is affected by relatively small radiation exposures, as is shown in Fig. 8. The dimethyl fluid-based compound hardens rapidly and then gels after a dose of approximately 25 megarads. The phenylmethyl fluid-based compound behaves in a similar manner except a larger dose is required to achieve the same effects.

Table II. Breakdown Voltage

		Average Breakdown Voltage, Kv	
Silicone Fluid, Cs	Previous Dose, Megarads	0.00 Megarads/ Hour	0.17 Megarad/ Hour
100, dimethyl	0	11.0	12.0
	50	8.3	11.8
40, phenylmethyl	2	12.0	11.8
	100	7.7	12.0

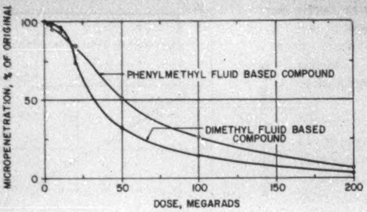


Fig. 8 (left). Consistency of two silicone compounds after irradiation

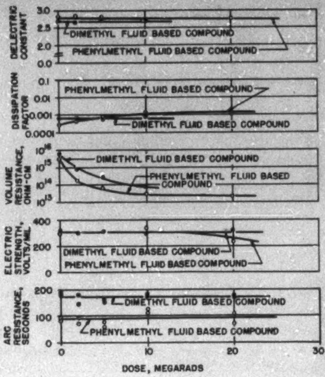


Fig. 9 (right). Dielectric properties of silicone compounds after irradiation

Gelation takes place after a dose of approximately 40 megarads.

After a dose of 200 megarads, both compounds become brittle gels. However, there is no visible evidence of shrinking or cracking even after 1,000 megarads of irradiation.

Similar to the phenylmethyl silicone fluids, the phenylmethyl fluid-based compound becomes amber colored upon radiation exposure. This is more apparent at the surface of the compound which is in contact with air. Apparently the discoloring reaction is between the phenyl groups in the compound and air. No change in color takes place in the dimethyl fluid-based compound even after 1,000 megarads of irradiation.

Dielectric Properties

The permanent effects of gamma radiation on the dielectric properties of silicone compounds are similar to those of silicone fluids. These effects are shown in Fig. 9.

Radiation doses less than the gelation dose have no appreciable effect upon the dielectric constant of either of these two silicone compounds. This radiation also has no effect no the dissipation factor of the dimethyl fluid-based compound; however, it causes a slight increase in the dissipation factor at 100 cps for the phenylmethyl fluid-based compound.

Radiation decreases the resistivity of both types of compounds. The resistivity of the phenylmethyl fluid-based compound was decreased a hundred fold by a dose of 10 megarads.

Fig. 9 shows that radiation has little effect on the electric strength. However, data in Table III for doses greater than the gelation dose show that the electric strength is sharply reduced by greater irradiation doses. The phenylmethyl fluid-based compound is more resistant in this regard; a dose greater than 200 megarads is required to achieve this decrease in electric strength.

The drop in electric strength by irradiation beyond gelation indicates that the physical integrity of the compound may be affected. This is not evident from visual observation since the defects are apparently microscopic in size.

Both Fig. 9 and Table III show that irradiation has little effect on the arc resistance of either silicone compound.

No measurement of the dielectric properties of silicone compounds has been made during radiation exposure. However, in view of the similarity in permanent effects of radiation on silicone compounds and fluids, it is expected that the transitory dielectric properties, which occur during irradiation, may also be similar.

Radiation Effects on Elastomers

Silicone elastomers are produced by introducing substantial amounts of inorganic fillers into high molecular weight silicone fluids and then vulcanizing, or cross-linking the polymer.[13] These elastomers are noted for retention of physical and dielectric properties during and after high- and low-temperature exposures. The primary effects of gamma radiation on silicone elastomers are demonstrated by data on five stocks in Table IV.

Physical Properties

Gamma radiation at 25 C hardens silicone elastomers and reduces the flexibility and tensile strength; Fig. 10 shows these effects for one silicone elastomer. These data are representative of a majority of the silicone elastomers. Elongation is the property most affected. It is reduced by 80% after 100 megarads of irradiation. After 500 megarads, the physical properties are essentially de-

Table III. Electric Strength and Arc Resistance of Irradiated Silicone Compounds

Dose, Megarads	Dimethyl Fluid-Based Compound: Gap, Inches	Dimethyl Fluid-Based Compound: Electric Strength, Volts/Mil	Dimethyl Fluid-Based Compound: Arc Resistance, Seconds	Phenylmethyl Fluid-Based Compound: Gap, Inches	Phenylmethyl Fluid-Based Compound: Electric Strength, Volts/Mil	Phenylmethyl Fluid-Based Compound: Arc Resistance, Seconds
0	0.050	316	183	0.050	320	86
2	0.050	306	163	0.050	316	88
5	0.050	308	160	0.050	292	70
10	0.050	296	183	0.050	328	112
20	0.050	320	173	0.050	226	92
50	0.050	256	160	0.050	248	58
100	0.132	284	183	0.161	242	121
200	0.200	98	189	0.163	251	116
500	0.215	79	172	0.140	106	154

Electric strength test: ASTM *D149-55T*, 0.5 kv per second rise.
Arc resistance test: *LP-406A* (ASTM *D495-56T*) (except pointed electrodes, 0.050-inch gap).

Table IV. Data on Silicone Elastomers

Elastomer (Silastic)	Polymer, Silicone	Typical Dielectric Application
80	Dimethyl	Molded parts, wire insulation
S-2048	Phenylmethyl	Molded parts
132	Dimethyl	Cloth coating
S-2007	Dimethyl	Heat vulcanized encapsulant
RTV S-5313, S-5314	Dimethyl	Room temperature vulcanized encapsulant

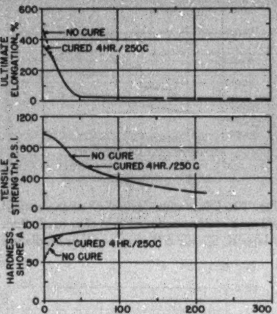


Fig. 10. Some physical properties of a typical silicone elastomer after irradiation. Material, Silastic 80

stroyed for most dielectric applications. Phenylmethyl-type silicone elastomers, such as Silastic* *S-2048*, have no marked improvement in resistance to these radiation effects.

Although the physical properties of silicone elastomers are severely damaged by doses as low as 50 megarads, many applications for these materials as dielectrics do not require extreme elongation and tensile strength. In such applications as coil insulation and encapsulants for electronic assemblies, silicone elastomers may be usable after doses as great as 200 to 500 megarads at 25 C.

Radiation has a permanent effect on the density of Silastic *RTV, S-5313, S-5314*. The density is increased by approximately 6% as the material shrinks after 100 megarads exposure. This may be a severe effect since the tensile strength is also damaged by irradiation. This large effect of radiation on density has not been observed for any other silicone elastomer.

DIELECTRIC PROPERTIES

Gamma irradiation produces both permanent and transient effects on the dielectric properties of silicone elastomers.

Permanent Effects. Gamma irradiation increases the dielectric constant of silicone elastomers. Fig. 11(A) shows this increase is largely a linear function of the irradiation dose and thus the dielectric constant of irradiated silicone elastomers can be expressed mathematically by equation 4:

$$K = K_0 + \beta_k D \qquad (5)$$

Values of β_k, the coefficient of increase in dielectric constant, have been calculated for frequencies of 100 cps and 10 mc/sec and temperature of 23 C.

Elastomer (Silastic)	β_k, Megarad^{-1}	
	100 Cps	10 Mc/Sec
Silastic *80*	0.00075	0.00057
Silastic *S-2048*	0.0000	0.0000
Silastic *S-2008*	0.0025	0.00093

Table V. Values of β_δ

Stock (Silastic)	β_δ, Megarad^{-1}	
	100 Cps	10 Mc
80	0.0015±0.0004	0.0015±0.0004
S-2048	0.0003±0.0002	0.0007±0.0002
S-2007	0.0025±0.0004	0.0011±0.0004
RTV S-5313, *S-5314* (after 40 megarads)	0.0000±0.0003	0.0000±0.0003

These data indicate that the radiation effects are more pronounced at 100 cps. However, regardless of the type of silicone elastomer, a dose of at least 200 megarads is required to increase the dielectric constant by 0.5; hence, for most insulating applications, radiation exposure has little effect on the dielectric constant of silicone elastomers.

Radiation exposure also increases the dissipation factor. Fig. 11(B) shows that the increase in dissipation factor can be expressed mathematically

$$\log\left(\frac{\tan\delta}{\tan\delta_0}\right) = \beta_\delta D \qquad (6)$$

when $\tan\delta$ is the dissipation factor, $\tan\delta_0$ is the dissipation factor prior to irradiation, β_δ is the coefficient of increase in dissipation factor, and D is the irradiation dose.

Values of β_δ at 100 cps and 10 mc/sec at 23 C temperature are shown in Table V.

These data indicate that considerable

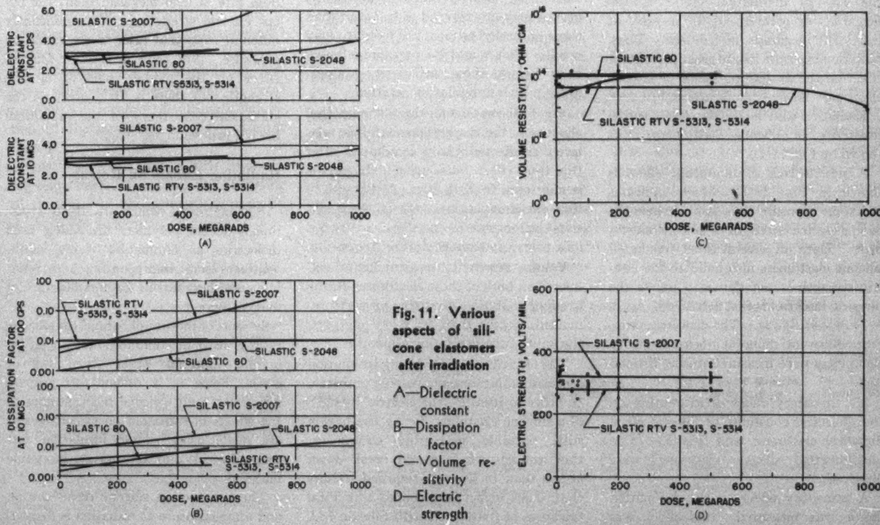


Fig. 11. Various aspects of silicone elastomers after irradiation

A—Dielectric constant
B—Dissipation factor
C—Volume resistivity
D—Electric strength

---

* Silastic is a registered trade name for Dow Corning Corporation silicone rubber.

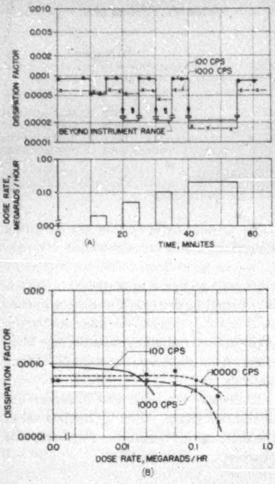


Fig. 12. Dissipation factors of dimethyl silicone elastomer. Material, Silastic 80, no previous irradiation

A—During and after irradiation
B—During irradiation

variation in radiation resistance exists among different silicone elastomers. However, for the silicone elastomers investigated, the dose required to double the 100-cps dissipation factor is 120 megarads or greater; at 10 mc/sec, at least 200 megarads is required. Thus, radiation exposure has no great permanent effect on the dissipation factor of silicone elastomers for most applications.

Radiation also has little effect on the resistivity of silicone elastomers as is shown by Fig. 11(C).

A similar lack of significant effect is shown in Fig. 11(D) which indicates radiation exposure does not decrease the 60-cps electric strength of silicone elastomers. Data on several other irradiated silicone elastomers also indicate the electric strength is not damaged unless the physical integrity is first destroyed.

Transient Effects. The dielectric properties of several different stocks of silicone elastomers were measured during irradiation.

No transitory effect of irradiation on the dielectric constants of Silastic *80*, a dimethyl elastomer, and Silastic *675*, a phenylmethyl silicone elastomer, were observed.

A transitory effect on the dissipation factor was measured; this effect was opposite in direction to that observed with silicone fluids. Fig. 12(A) shows that the dissipation factor of a dimethyl silicone elastomer decreases with increasing radiation intensity. This effect is pronounced at 100 cps but decreases with increasing frequencies and is not easily observed at 10,000 cps at dose rates of 0.1 megarad/hour or less.

No explanation for this effect is offered except that by some mechanism the absorbed radiation energy apparently reduces either the charge or mobility of large charged regions of the material.

Fig. 12(B) shows that the dissipation factor of this silicone elastomer during irradiation does not follow a simple relationship as was found for silicone fluids.

Similar measurements were also made on Silastic *675*, a phenylmethyl silicone elastomer. Data obtained show that in addition to the general effects on a dimethyl elastomer, there is a time-dependent effect both during and immediately after irradiation. As indicated in Fig. 13, the magnitude of this time-dependent effect during irradiation is essentially the same as that immediately after irradiation. The time constant of this effect is approximately 6.5 minutes for irradiation exposures of 10 to 20 minutes at dose rates from 0.02 to 0.2 megarad/hour.

It is suggested that this effect is due to products of radiation-induced reactions involving the phenyl groups of the silicone elastomer. Since products of the phenyl groups may be expected to be larger than those of methyl groups, less mobility may result. This would then require a longer time before these activated products achieve their annihilating reaction.

Fig. 14 shows that for this phenylmethyl elastomer, the dissipation factor is relatively unaffected during irradiation after this time effect takes place. It can be readily seen in both Figs. 12(B) and 14 that the dissipation factor of these silicone elastomers is essentially not damaged in a transient manner during irradiation.

Volume resistivity measurements conducted on both of these elastomers during irradiation showed that this property remained greater than 0.5×10^{14} ohm-cm, the limit of the measuring apparatus.

The electric strength during irradiation is also an important engineering property. In order to measure the electric strength of a silicone elastomer during irradiation with portable measuring equipment, thin samples which would break down at less than 14 kv were required. Glass cloth 0.004 inch thick, coated to a total thickness of 0.010 inch with Silastic *132*, a cloth-coating silicone elastomer, was used for this study.

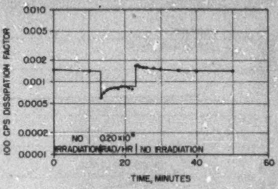


Fig. 13. Dissipation factors of phenylmethyl silicone elastomer during and after irradiation. Material, Silastic 675; no previous irradiation

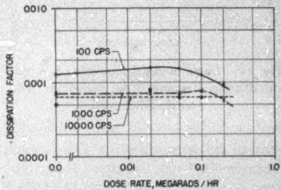


Fig. 14. Dissipation factor of phenylmethyl silicone elastomer during irradiation. Material, Silastic 675; 5 minutes previous irradiation

Electric strength data often follow an extreme value statistical distribution.[14,15] Data from these measurements were therefore plotted on extreme value probability paper. Fig. 15 shows that at a dose rate of 0.06 megarad/hour, the 60-cps electric strength is not significantly different from that prior to the radiation exposure. Not only is there a lack of effect on the modal value of electric strength, but there is little effect on the statistical consistency of the individual breakdown voltages.

Radiation Effects on Resins

Silicone resins, which are highly cross-linked silicone polymers, are widely used dielectrics as laminating resins, cloth-coating resins, impregnating and bonding varnishes, and in the manufacture of magnet wire.[16]

In most of these applications the silicone resin is used in combination with an inorganic material, often glass fiber or cloth. Since it is difficult to prepare samples of pure silicone resin, studies of the effects of radiation on silicone resins are made using simple insulation systems of resins and appropriate inorganic materials.

The behavior of silicone resins during and after exposure to radiation is demon-

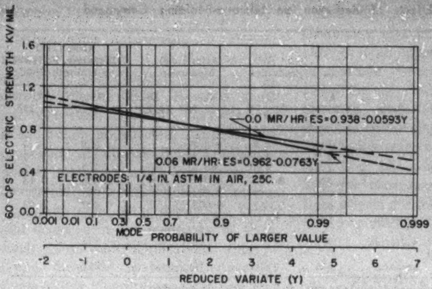


Fig. 15. Electric strength of silicone elastomer coated glass cloth before and during irradiation. 0.004-inch cloth coated with Silastic 132 to 0.010 inch thickness

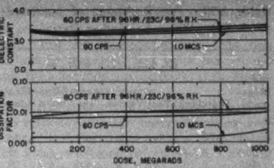


Fig. 16. Dielectric constant and dissipation factor of silica-filled silicone resin after irradiation

strated by data on four simple resinous insulating systems:

Resin	Dow Corning	Inorganic Material
Solventless resin	*R-7521*	Silica flour
Cloth-coating resin	*994* (varnish)	Glass cloth
Molding compound	*301*	Chopped glass fiber
Laminating resin	*2105 A*	Glass cloth

PHYSICAL PROPERTIES

Measurements on irradiated samples of silica-filled solventless silicone resin provide no evidence that irradiation doses up to 1,000 megarads affect the flexural strength. These data are listed in Table VI.

The only salient effect of radiation exposure on the physical properties of these resins is a decrease in flexibility of the glass cloth-resin systems after 500 and 1,000 megarads of irradiation. This is evidence of extensive radiation-induced cross-linking of the resin. However, the irradiated glass cloth insulating systems were easily flexed over a small mandrel which shows the ultimate elongation of the irradiated resins is greater than 2%.

One other effect, but of little engineering importance, is a noticeable darkening of the color of the solventless resin by doses of 500 megarads and greater.

Table VI. Flexural Strength of Irradiated Silica-Filled Solventless Resin

Dose, Megarads	Flexural Strength, Psi				Average Flexural Strength, Psi
0	8,940	8,100	8,400		8,480
200	8,580	7,650	7,970		8,067
500	8,700	7,780	8,720		8,400
1,000	8,100	8,220	8,200	7,940	8,115
				Average:	8,254±260

Test: ASTM *D790-49T*.
Subject: Approximately 4×1×1/8 inch slabs.

DIELECTRIC PROPERTIES

Similar to the effects on silicone elastomers, radiation has some, but few permanent, effects on the dielectric properties of silicone resin systems. The transient effects, however, are similar to those present with silicone elastomers.

Permanent Effects. Irradiation causes a slight increase in the dielectric constant of silicone resins as demonstrated by Fig. 16 for the solventless resin. Even after a dose of 1,000 megarads the increase in dielectric constant of silica-filled solventless resin at frequencies from 60 cps to 100 mc is so small as to be of little engineering significance. This figure also shows that the dielectric constant of irradiated solventless silicone resin is little affected by additional moisture exposure. This indicates irradiation does not greatly affect the water adsorption characteristics of the volume of the silicone resin.

An irradiation dose of 500 megarads has no appreciable effect on the dissipation factor of the solventless silicone resin at frequencies from 100 cps to 10 mc. However, as is also shown in Fig. 16, irradiation from 500 to 1,000 megarads increases the dissipation factor; after 1,000 megarads the 60-cps dissipation factor is approximately 0.01.

Fig. 17 shows that irradiation has no appreciable effect on the electric strength of the solventless silicone resin unless the dose exceeds 500 megarads. An additional 500 megarads, however, increases the rapid-rise electric strength by more than 60%; the 1-kv-per-minute step-by-step electric strength is also increased about 25%. The rate of voltage rise has no appreciable effect on the electric strength for irradiation doses of 500 megarads and less. Thus, irradiation beyond 500 megarads affects the voltage endurance relationship.

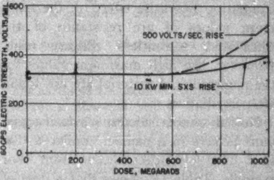


Fig. 17. Electric strength of silica-filled silicone resin after irradiation

Data given in Table VII, for this resin system after 1,000 megarads of irradiation, indicate that the electric strength with a slower rate of voltage rise would be less than with a 1-kv-per-minute rise. Subsequent data for 1-kv-per-10-minute rate of voltage rise show that this is true; however, electric strength under these conditions is still greater than the strength of the samples irradiated 500 megarads. Thus, although irradiation affects the electric strength of a silicone resin system, the effects are not damaging.

Table VII. Breakdown Voltage of Irradiated Silica-Filled Solventless Resin

		Breakdown Voltage, Kv								
Dose, Megarads	Thickness, Inches	500 Volts per Second	1 Kv per Minute, Step by Step			1 Kv per 10 Minutes, Step by Step				Electric Strength, Volts per Mil for 1 Kv per Minute Rise
0	130	40	43	40	43	41				322
200	122	42	41	43	40	40				336
500	133	43	43	43	40	40				312
1,000	123	64	49	46	55	46	46	44	46	398

Test: 60 cps with 1/4-inch ASTM electrodes in oil at 23 C, ASTM *D149-55T*.
Subject: Slabs approximately 4×4×1/8 inch.

Table VIII. Arc Resistance of Irradiated Silica-Filled Solventless Resin

Dose, Megarads	Arc Resistance, Seconds: Individual Test Values					Average	Minimum	Maximum
0	122	4	104	8	6	49	4	122
200	6	4	6	8	4	5.6	4	8
500	6	6	4	4	4	4.8	4	6
1,000	6	8	8	8	7	7.4	7	8

Test: ASTM *D495-56T*.
Subject: Same as used for electric strength measurements; arc resistance measurements made after electric strength tests.

Table IX. Effects of Radiation on Silicone-Molding Compound

Property	Condition	Units	Prior to Irradiation A-24/23/50*	Prior to Irradiation D-24/23*	After 1,000 Megarads A-24/23/50*	After 1,000 Megarads D-24/23*	After 1,000 Megarads Followed by 1 Month/300C A-24/23/50*	After 1,000 Megarads Followed by 1 Month/300C D-24/23*
Dielectric constant	100 cps		4.3	4.8	4.7	†	3.8	7.7
	1,000 cps		4.2	4.7	4.6	†	3.7	7.1
	1.0 Mc/sec		4.1	4.3	4.4	†	3.7	4.9
Dissipation factor	100 cps		0.006	0.016	0.02	†	0.006	0.18
	1,000 cps		0.006	0.017	0.014	†	0.005	0.11
	1.0 Mc/sec		0.003	0.012	0.005	†	0.002	0.08
Volume resistivity	1.0 min. electrification	ohm-cm	1.5×10^{14}	0.5×10^{14}	0.6×10^{14}	10^{8}	0.3×10^{14}	1.5×10^{10}
Electric strength	1/4 inch electrodes in oil	volts/mil	271		69		142	
Arc resistance		seconds	235		76		247	

* A-24/23/50: Sample conditioned 24 hours at 23 C and 50% relative humidity.
D-24/23: Sample conditioned 24 hours in distilled water at 23 C.

† Beyond range of measuring techniques used.

The data in Table VIII show that irradiation has little effect on the minimum values of arc resistance of the silica-filled solventless silicone resin. However, the high maximum values for the nonirradiated material are not found after irradiation.

Not all simple silicone insulating systems behave in a manner similar to the silica-filled solventless silicone resin. For example, Table IX shows that a dose of 1,000 megarads has a sizeable effect on the dielectric constant, dissipation factor volume resistivity, and electric strength of a silicone molding compound. Irradiation reduces the water resistance of this material. Much of this irradiation damage is removed, however, by subsequent high-temperature exposure. Table IX shows that a 1-month exposure at 300 C restores most of these properties.

Transient Effects. Dielectric constant dissipation factor, and volume resistivity measurements made on a silicone-glass laminate during and immediately after irradiation were similar to those made on silicone elastomers.

No transient effect on the dielectric constant was observed.

As shown in Fig. 18, the effect on dissipation factor is similar to that on a dimethyl silicone elastomer. Although the laminating resin contains phenyl groups, no time-dependent effect was observed. Similar to the elastomers, the decrease in dissipation factor during irradiation is more pronounced at the lower frequencies:

Frequency, Cps	Dissipation Factor: 0.0 Megarads/Hour	Dissipation Factor: 0.2 Megarads/Hour
100	0.0024	0.0004
1,000	0.0030	0.0012
10,000	0.0013	0.0014

During irradiation at dose rates up to 0.2 megarad/hour, the volume resistivity remained greater than 0.3×10^{14} ohm-cm, the limit of the measuring system and slightly less than the resistivity prior to the radiation exposure.

The electric strength of silicone resins during irradiation is also an important property. To measure this characteristic, 0.004-inch-thick glass cloth coated with a silicone cloth coating resin was used. Data from these measurements were plotted in Fig. 19 on extreme value probability paper. These data show that dose rates up to 0.15 megarad/hour have no appreciable effect on the modal value of the electric strength. Furthermore these data show a possible increase in the consistency of the breakdown voltages with increase in radiation intensity.

Conclusions

Irradiation ionizes silicone dielectrics causing rearrangement of the chemical structure and the production of impurities. This rearrangement of the chemical structure causes major effects on the physical properties.

The rearrangement and the impurities produced have significant effects on some of the dielectric properties. Since some of the impurities are transitory, some of the effects are also transitory, while others are permanent. These transitory effects in silicone dielectrics are often larger than the permanent effects and occur during and for some silicones immediately after irradiation. The transitory effects may limit the performance of silicone fluids as dielectrics during radiation exposure.

In general, the physical properties of silicone fluids, compounds, elastomers, and resins are damaged by irradiation before the essential dielectric properties are significantly and permanently affected. Therefore, the effects of irradiation upon the physical properties are the primary factors in limiting the performance of silicone dielectrics after radiation exposures.

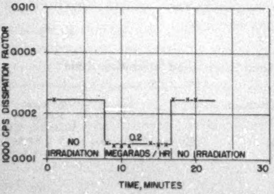
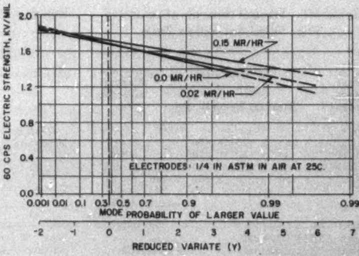


Fig. 18. Dissipation factor of silicone-glass laminate during and after irradiation. Sample received no previous irradiation

Fig. 19. Electric strength of silicone resin coated glass cloth before and during irradiation. Dow Corning 936 Resin coated on 0.004-inch glass cloth to 0.008-inch thickness



Silicone Fluids

Other than a decrease in electric strength, radiation at 25 C causes few effects on the dielectric properties of dimethyl silicone fluids. However, physical properties, primarily viscosity and gas evolution, are affected. The physical properties of silicone fluids containing phenyl groups are much less affected by irradiation; however, irradiation causes a drastic increase in the dissipation factor and decrease in the resistivity of these fluids. For those applications where a liquid dielectric is required dimethyl fluids of low viscosity are most practical and may be usable for doses up to 100 megarads.

Silicone Elastomers

Irradiation doses less than 200 megarads have little significant effect on the dielectric properties of silicone elastomers. The physical properties of silicone elastomers, however, are affected by relatively low doses. After a dose of 100 megarads, a majority of the flexibility of the elastomer is destroyed. For those numerous applications in electrical and electronic apparatus where extreme flexibility is not required once the elastomer is in place, silicones may be suitable for doses of 200 to 500 megarads.

Silicone Resins

Silicone resins, in general, are much more resistant to radiation than silicone fluids and elastomers. The major dielectric properties of a solventless silicone resin are not damaged by gamma irradiation doses as high as 1,000 megarads. These doses also have no significant effect on the physical integrity or strength of this material. Although the radiation resistance of this material is believed to be typical of those on most silicone resins, radiation has been found to destroy important dielectric properties of one silicone molding compound. These dielectric properties, however, are restored to a considerable degree by subsequent high-temperature exposure.

Dielectric Properties During Irradiation

The dielectric constant of all silicone dielectrics tested remains essentially the same during irradiation as before. Electric strength is likewise unaffected.

The electrical losses are affected, however. The dissipation factor of silicone fluids at 100 cps and 1,000 cps may be 10 to 100 times normal when irradiated at a dose rate of 0.1 megarad/hour. The resistivity is likewise affected.

The dissipation factor of silicone elastomers and at least one resin are affected in an opposite manner, however. With these dielectrics, the dissipation factor during irradiation decreases with increasing radiation intensity at frequencies below 10,000 cps.

Although some time-dependent effects have been observed with silicone fluids and one elastomer, the magnitudes of these effects are relatively small and the time constant less than 10 minutes.

Appendix. Samples and Test Methods

Unless otherwise noted, all radiation exposures are made at 25±3 C and at a dose rate of 0.5 to 0.7 megarad/hour. Measurements are in general made at 23±1 C in accordance with applicable ASTM (American Society for Testing Materials) standards:

Property	ASTM Standard Test Method
Dielectric constant	D150-54T
Dissipation factor	D150-54T
Volume resistivity	D257-57T
Electric strength	D149-55T
Arc resistance	D495-56T
Flexural strength	D790-49T

Fluid samples consist of 3 ounces of fluid in 4-ounce bottles vented through small holes in the caps. These samples are irradiated at a dose rate of 0.3 megarad/hour. Compounds are placed in 2-ounce tin ointment cans, $2^3/_8$ inches diameter by 3/4 inch deep, and cover applied.

Elastomers are press molded into sheets approximately 0.10 inch thick and cut into slabs $3^1/_2 \times 8$ inches. Resin systems are cast or molded into samples approximately 0.125 inch thick and 3 to 10 square inches in area. The resin samples are wrapped in 0.006-inch-thick aluminum foil prior to radiation exposure.

Measurements After Irradiation

Dielectric constant, dissipation factor, and volume resistivity measurements are made using the same cell: fluids, J. C. Balsbaugh type *2TN25* cell (2-element); compounds, 2-element cell with vacuum capacitance of 35 micromicrofarads; elastomers and resins, the General Radio Company type *1690-A* Sample Holder (2-element). Measurements are made at 50, 10^2, 10^3, 10^4, 10^5 cps using General Radio type *716-C* Schering Bridge and appropriate auxiliary apparatus. These measurements are also made at 150 C on the fluid samples. Measurements are made at 1 and 10 megacycles/second using a susceptance variation circuit, and in the case of fluids and compounds, a special 2-element cell. Resistivity measurements are made using a Federal Telephone and Radio type *H* Terraohmmeter. An electrification period of 1 minute is used and the test potential is 500 volts.

Electric strength of irradiated fluids and compounds are made using 0.500-inch-diameter spherical steel electrodes. For fluids, the d-c test voltage is applied step-wise 1.0 kv per minute and the electrode gap is 0.020 inch. For the compounds, the cps test voltage is increased 500 volts per second and the electrode gap is 0.050 inch except for the gelled specimens.

Electric strength of irradiated elastomer and resin samples are measured using 1/4-inch ASTM electrodes on oil. The 60-cps test voltage is increased step-wise 1 kv per minute except where otherwise noted with these data.

Arc resistance is measured using pointed electrodes for the compounds and elliptical electrodes for the resin samples. The electrode gap is 0.50 inch.

The viscosity of irradiated fluids is measured using a capillary viscosity pipette or a Brookfield model *LVF* viscometer. The density is measured using Fisher hydrometers.

The consistency of irradiated compounds is measured using a precision scientific micropenetrometer.

The hardness of elastomer samples is measured on a Shore *A* scale. The ultimate elongation and tensile strength are measured using an Instron model *TM* tensile tester.

The flexural strength of irradiated solventless resin samples are measured using a Baldwin Universal Tester.

Measurements During Irradiation

Dielectric constant, dissipation factor, resistivity, and electric strength of fluid, elastomer, and resin samples are also measured during irradiation.

A J. C. Balsbaugh *3TN25* 3-terminal cell is used for complex permittivity and resistivity measurements of silicone fluids. Elastomer and resin samples, 2 inches in diameter, are placed in a guarded 3-terminal cell with thin aluminum electrodes having an active surface of 1.500 inches in diameter. These cells and the lead connections are completely submerged in Dow Corning 200 fluid, 250 cs during the measurements. The elastomer and resin samples are wetted with this silicone fluid to insure an absence of air pockets next to the electrodes. Guarded leads are used between the measuring instruments outside of the "cave" and the electrodes. Dielectric constant and dissipation factor are measured using an Electro-Measurements model *270* capacitance bridge at 100, 1,000, and 10,000 cps; resistivity is measured using a General Radio model *544-B* megohm bridge.

The electric strength of fluids during irradiation is measured by the same procedure as detailed for irradiated fluids after irradiation. The electric strength of elastomer and resin coated glass cloth is measured using 1/4-inch ASTM electrodes in air. The 60-cps test voltage is increased at a rate of 200 to 500 volts per second.

The shielding effect of the sample holders is considered in the orientation of the sample and in the calculation of the dose rate.

References

1. The Effects of High-Energy Gamma Radiation on Dielectric Solids, Philipp H. Klein, Clifford Mannal. *AIEE Transactions*, vol. 74, pt. I (*Communication and Electronics*), 1955 (Jan. 1956 section), pp. 723–29.

2. Basic Concepts of Radiation Chemistry

Applicable to Effects of Radiation on Dielectrics, *Report*, Conference on Effects of Radiation on Dielectric Materials, **M. Burton.** *OTS PB 111863*, U. S. Department of Commerce, Washington, D. C., Dec. 1954, pp. 1–12.

3. **Beneficial Effects of Radiation on Polymers, Arthur Charlesby.** *Nucleonics*, New York, N. Y., vol. 14, no. 9, Sept. 1956, pp. 82–85.

4. **Conductivity Changes in Dielectric During 2.5 Mev X-Radiation, F. C. Armistead, J. C. Pennock, L. W. Mead.** *Physical Review*, New York, N. Y., vol. 76, 1949, p. 860.

5. **Chart of the Nuclides** (Radiological Health Handbook), **G. Friedlander, M. Perlman, J. R. Stehn, E. F. Clancy.** *OTS PB 121784*, U. S. Department of Commerce, Jan. 1957, p. 88.

6. **Radiation Damage to Elastomers, Plastics, and Organic Liquids, C. G. Collins, U. P. Calkins.** *APEX 261*, U. S. Department of Commerce, Sept. 1956, pp. 11–12.

7. **Silicones and Their Uses, R. R. McGregor.** McGraw-Hill Book Company, Inc., New York, N. Y., 1954, pp. 36–80.

8. **Beneficial Effects of Radiation on Polymers, Arthur Charlesby.** *Nucleonics*, vol. 14, no. 9, Sept. 1956, p. 86.

9. **Melt Viscosities of Dimethylpolysiloxanes, E. L. Warrick, W. A. Piccoli, F. O. Stark.** *Journal*, American Chemical Society, Washington, D. C., vol. 77, 1955, p. 5017.

10. **Viscometric Investigation of Dimethylsiloxane Polymers, A. J. Barry.** *Journal of Applied Physics*, New York, N. Y., vol. 17, 1946, p. 1024 ff.

11. **Effects of Radiation on Organopolysiloxanes, E. L. Warrick.** *Industrial and Engineering Chemistry*, Washington, D. C., vol. 47, Nov. 1955, p. 2389 ff.

12. **Silicones and Their Uses, R. R. McGregor.** McGraw-Hill Book Company, Inc., 1954, pp. 87, 93.

13. **Silicones and Their Uses, R. R. McGregor.** McGraw-Hill Book Company, Inc., 1954, pp. 149–86.

14. **Electrode Area Effect for the Impulse Breakdown of Transformer Oil, H. S. Endicott, K. N. Weber.** *AIEE Transactions*, vol. 76, pt. III (*Power Apparatus and Systems*), Aug. 1957, pp. 393–98.

15. **Impulse Breakdown of Electrically Prestressed Polyethylene, C. R. Vail, W. F. Gauster.** *Ibid.*, Apr. 1957, pp. 38–41.

16. **Silicones and Their Uses, R. R. McGregor.** McGraw-Hill Book Company, Inc., 1954, pp 108–49.

Mathematics of Insulation-Aging Calculations

L. C. WHITMAN
FELLOW AIEE

W. W. WHITMAN
NONMEMBER AIEE

Synopsis: Equations are developed for the calculation of insulation aging in apparatus, in which it is recognized that the temperature of the apparatus usually changes in accordance with an exponential relation. The results from this method are compared with the simplification usually made of assuming rectangular temperature cycles and also with the assumption of a triangular temperature cycle. Insulation aging rate is based on the use of the Arrhenius chemical reaction rate theory, which relates insulation deterioration to the reciprocal of the absolute temperature. All aging values are in terms of relative aging, which assigns a value of unit aging to a specified temperature. To establish the errors involved in the simplifying assumptions conventionally used in aging computations, the calculations in this paper are carried out to more significant figures than basic aging data and time constants would justify.

THE TRANSIENT temperature of most electric apparatus, for the condition of increasing temperature, can usually be quite accurately expressed as an exponential function of time[1,2] as follows (Appendix I gives the nomenclature):

$$T=T_0'+A\,(1+\epsilon^{-\alpha t}) \qquad (1)$$

Also in the cooling of electrical apparatus the transient temperature can be similarly expressed as

$$\begin{aligned} T&=T_1-D\,(1-\epsilon^{-\alpha t})\\ &=T_0'+D\epsilon^{-\alpha t} \qquad (2)\end{aligned}$$

A typical example of the transient temperature of electric apparatus through one load cycle is shown in Fig. 1, assuming a load $=L_1$ from t_0 to t_1 and a load L_0 from t_1 to t_n. L_0 is here the steady-state load existing for a length of time previous to t_0 to establish an equilibrium temperature condition at t_0.

The examples given in this paper assume that the time constant is the same in equations 1 and 2. This is approximately true for air-cooled apparatus where there is a single thermal gradient from the apparatus to the cooling medium (air). In liquid-cooled apparatus this is not so since there are essentially two thermal gradients in series (apparatus to liquid and liquid to air) with quite different time constants.

Under transient temperature conditions insulation-aging rates are continuously changing throughout the load cycle. Simplifying assumptions concerning temperature are usually made to simplify these aging calculations. The simplest assumption (illustrated in Fig. 2) is that the shaded portions of Fig. 1 (or at least their effects) are equal. For the case shown the shaded areas differ by some 67%, and still differ by about 8% for the case of $t_1=10$ hours, becoming equal only when both t_1-t_0 and $t_{24}-t_1$ are infinite.

The insulation aging from time t_0 to t_{24} under the conditions of Fig. 1 is also quite different from that shown for the same time interval in Fig. 2. It will later be shown that this assumption leads to answers too large for large time intervals t_0 to t_1 and to answers too small for shorter time intervals.

An alternate assumption is shown in Fig. 3. In this case the answers for insulation aging will invariably be too low, irrespective of the length of time that L_1 exists, or of apparatus time constant. The magnitude of this error will be later shown for some illustrative cases.

Relative Insulation Aging

It is generally recognized that most insulations deteriorate in accordance with the chemical reaction rate theory.[3] From this it follows that deterioration rates of insulation are generally straight lines when the reciprocal of the absolute temperature of the insulation are plotted as abscissas and logarithms of life in time unit (hours or years) are plotted as ordinates. It is difficult to obtain data or to obtain agreement regarding the end of life of insulation. To make this study perfectly general it is thus convenient to use a relative insulation-aging curve instead of an absolute one. Such an approach requires only knowledge of the aging rate for a particular insulation and the assignment of unit aging to a definite reference temperature. Representative relative aging rate curves are given for two different types of insulation for two different reference temperatures together with data for the rates used in determining the slopes of these curves (given in references 4 and 5). The reference temperatures are 10 degrees lower than the maximum hottest spot temperatures allowed for these classes of insulations, corresponding to an average ambient of 30 C (degrees centigrade) over 24 hours, rather than a maximum

---

Paper 59-118, recommended by the AIEE Electrical Insulation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 29, 1958; made available for printing December 16, 1958.

L. C. Whitman is with the General Electric Company, Pittsfield, Mass., and W. W. Whitman is with Cornell University, Ithaca, N. Y.

Vol.
Pt. 1
1959

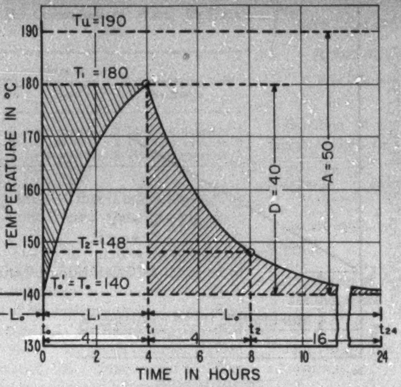


Fig. 1. Typical example of transient temperature of electric apparatus: case 3 in calculations

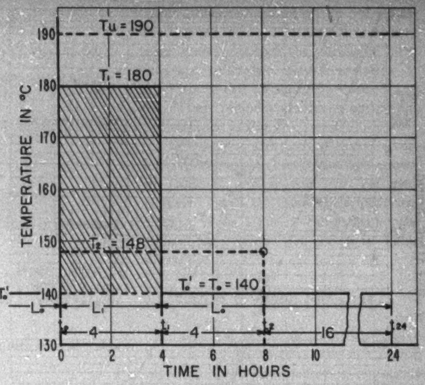


Fig. 2. A simplifying assumption of the transient temperature cycle of Fig. 1 often used in aging calculations. Case 1 in calculations

ambient of 40 C over 24 hours. These relative aging curves from reference 4 are shown in Fig. 4.

The equation for these curves is

$$\ln R = b\left[\frac{1}{T_0} - \frac{1}{T}\right] \qquad (3)$$

Equations will now be developed covering insulation aging for the cases of rectangular, triangular, and exponential temperature cycles and computations made illustrating the variation in results.

Rectangular Time-Temperature Cycle: Case 1, Fig. 2

Case 1 is quite simply calculated. As an illustrative example, assume

$T_0 = 140 + 273 = 413$ K (degrees Kelvin)
$T_1 = 180 + 273 = 453$ K
$t_1 - t_0 = 4$ hours
$t_{24} - t_1 = 20$ hours

Also assume relative aging curve *B* in Fig. 4, where $b = 11{,}500$; relative aging at 180 C from Fig. 4 = 12.

This may be more accurately determined by the use of equation 3, which gives the value 11.69 (with assumed values just given of temperature and $b = 11{,}500$). To compute cumulative relative aging over the 24-hour cycle assumed in Fig. 2 the relative aging at each temperature must be multiplied by the elapsed hours at the given temperature. Note that relative aging at the assumed reference temperature of 140 C equals 1 (from the definition of relative aging). The cumulative aging from t_1 to t_{24} then equals $(1)(20) = 20$.

Hence, hours' cumulative relative aging over the 24-hour temperature cycle in Fig. $2 = (4)(11.69) + 20(1) = 66.8$. The reference aging for 24 hours is obviously $(24)(1) = 24$.

$$\text{Relative aging over 24-hour temperature cycle} = 66.8/24 = 2.78 \qquad (4)$$

This means that if such a temperature cycle is repeated daily the insulation system will have 1/2.78 of the hours of life it would have with a continuous temperature of 140 C.

Triangular Time-Temperature Cycle: Case 2, Fig. 3

Case 2 would appear to be a simple calculation but it is actually quite involved since it requires integration under a relative aging curve. A general solution for this case is given in Appendix II, as shown in equation 40.

This general solution involves exponential integrals for which tables are available[6] for values of *x* between between zero and 15, and[7] between 20 and 50, with a region for *x* between 15 and 20 with no available tables. A simplified expression may be used which has an error of less than 1.0% for $x > 10$ (*x* is expressed in equation 40 by *u*) with the error decreasing to less than 0.05% for $x > 20$. In the particular example (and in most cases encountered) the limits *u* and u_0 on the integral are greater than 20, so the simplified equation 43 in Appendix II is applicable. Also, in Fig. 3, $T_0' = T_0$, so that the more general case given by equation 43 (as

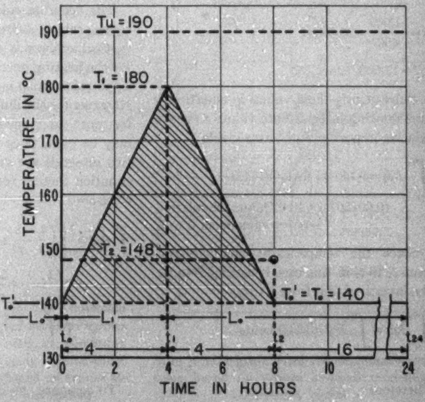
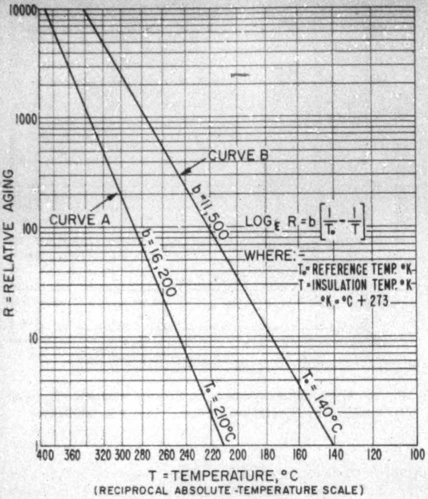


Fig. 3 Alternate simplifying assumption of the transient temperature cycle of Fig. 1 which might be used in aging calculation. Case 2 in calculations



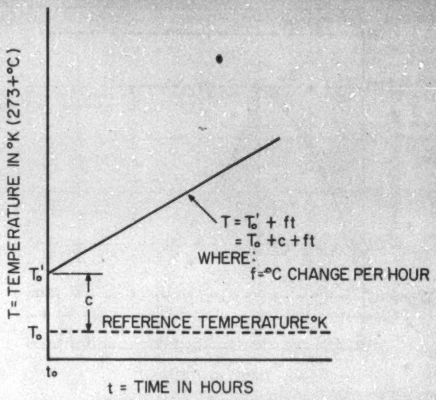


Fig. 4 (left). Representative relative aging curves for two different types of insulation for two different reference temperatures

Fig. 5 (above). Time-temperature relations for a constant rate of temperature change; see Appendix II for insulation aging equations

derived from Fig. 5) reduces to equation 44 (Appendix II) which is

$$\int_0^t R\,dt = \frac{b}{f}\left[\epsilon^{u_0-u}\left(\frac{1}{u^2}-\frac{2}{u^3}+\frac{6}{u^4}\right)-\left(\frac{1}{u_0^2}-\frac{2}{u_0^3}+\frac{6}{u_0^4}\right)\right] \quad (5)$$

To make a comparison of cases 1 and 2 assume the same values of these quantities as in Case 1; then

$$f = \frac{T_1 - T_0}{t} = \frac{453-413}{4}$$

$= 10$ C per hour temperature change

$$u_0 = \frac{11{,}500}{413} = 27.84504$$

$$u = \frac{11{,}500}{453} = 25.38631$$

$b/f = 11{,}500/10 = 1{,}150$

Substituting these values in equation 5 and showing intermediate values to show relative importance of terms yields

$$\int_0^{t_1} R\,dt = 1{,}150[11.69(0.001445) - (0.001207)] = 18.02$$ hours cumulative relative aging from t_0 to t_1

Since the temperature characteristic from t_1 to t_2 in this case is assumed to be the inverse of that from t_0 to t_1,

$$\int_{t_1}^{t_2} R\,dt = \int_0^{t_1} R\,dt = 18.02 \text{ hours}$$

cumulative relative aging from t_1 to t_2

Therefore

$$\int_0^{t_2} R\,dt = 18.02 + 18.02 = 36.04 \text{ hours}$$

cumulative relative aging from t_0 to t_2

The cumulative relative aging from t_2 to $t_3 = (1)(16) = 16$ hours. Hence the cumulative relative aging over the 24-hour temperature cycle in Fig. 3 $= 36.04 + 16 = 52.04$ hours. Therefore

Relative aging over the 24-hour temperature cycle of Fig. 3 $= \frac{52.04}{24} = 2.17$ (6)

Exponential Time-Temperature Heating Cycle: Case 3a, Fig. 1

The accurate calculation of case 3a is considerably more involved. Mathematical analysis is given in Appendix III for the heating portion of the cycle where a general solution is shown in equation 78. However for the illustrative example (and for most cases encountered) equation 79 may be used, since the limits on the first two integrals are greater than 10. This equation further reduces to equation 80, since $c=0$.

$$\int_0^t R\,dt = \frac{1}{\alpha}\left[\left(\frac{1}{p}-\frac{1}{p^2}+\frac{2}{p^3}-\frac{6}{p^4}\right) - \epsilon^{p-s}\left(\frac{1}{s}-\frac{1}{s^2}+\frac{2}{s^3}-\frac{6}{s^4}\right) - \epsilon^{p-r}\left(\int_{p-r}^{\infty}\frac{\epsilon^{-y}}{y}\,dy - \int_{s-r}^{\infty}\frac{\epsilon^{-y}}{y}\,dy\right)\right] \quad (7)$$

To compare the insulation aging based on exponential temperature change with the examples given in cases 1 and 2, values of $\alpha = 0.4$ and $A = 50$ were selected so that for case 3 the value of T_1 at the end of $t_1 = 4$ hours agreed closely with the value of T_1 assumed in cases 1 and 2. These values with the previously assumed value of b and T_0 in cases 1 and 2 give

$$\frac{1}{\alpha} = 2.5$$

$$A = 50$$

$$T_1 = 452.905$$

$$p = \frac{11{,}500}{413} = 27.84504$$

$$r = \frac{11{,}500}{50+413} = 24.83801$$

$$s = \frac{11{,}500}{413+50(1-\epsilon^{-(0.4)(4)})} = 25.39164$$

Substituting these values in equation 7 and showing intermediate values to show the relative importance of terms yields

$$\int_0^{t_1} R\,dt = 2.5[(0.034726) - 11.6278(0.037940) - 20.230(0.012932 - 0.49957)]$$

$= 23.688$ hours cumulative relative aging from t_0 to t_1 (8)

Equation 7 contains exponential integrals of the form

$$-Ei(-x) = \int_x^{\infty}\frac{\epsilon^{-y}}{y}\,dy \quad (9)$$

These are evaluated by tables in references 6 and 7.

Table I. Comparison of Aging for Three Cases

Case No.	Fig. No.	Type of Temperature Cycle	Relative Aging Over 24 Hours
1	2	rectangular	2.78, from equation 4
2	3	triangular	2.17, from equation 6
3	1	exponential	2.412, from equation 13

Exponential Time-Temperature Cooling Cycle: Case 3b, Fig. 1

For case 3b, to complete the calculation of the aging of the insulation subjected to the temperature cycle in Fig. 1, use is made of the mathematical analysis given in Appendix IV for the cooling portion of the cycle, where a general solution is given in equation 110. However, for the illustrative example (and for cases usually encountered) equation 111 may be used since the limits on the first two integrals are greater than 10. This equation further reduces to equation 112, since $c=0$.

$$\int_{t_1}^{t_{24}} R\,dt=\frac{1}{\alpha}\left[\epsilon^{M-L}\left(\frac{1}{L}-\frac{1}{L^2}+\frac{2}{L^3}-\frac{6}{L^4}\right)-\epsilon^{M-N}\left(\frac{1}{N}-\frac{1}{N^2}+\frac{2}{N^3}-\frac{6}{N^4}\right)+\int_{-\infty}^{M-L}\frac{\epsilon^n dn}{n}-\int_{-\infty}^{M-N}\frac{\epsilon^n dn}{n}\right] \qquad (10)$$

With the previously assumed values of α, b, T_0, T_1, and t_1, as in cases 1, 2, and 3a the quantities give

$$\frac{1}{\alpha}=2.5$$

$$D=T_1-T_0=452.905-413.0=39.905$$

$$L=\frac{b}{T_1}=\frac{11{,}500}{452.905}=25.39164$$

$$M=\frac{b}{T_0}=\frac{11{,}500}{413}=27.84504$$

$$N=\frac{b}{T_0+D\epsilon^{-\alpha t}}=\frac{11{,}500}{413+39.905\epsilon^{(-0\cdot 4)(20)}}=27.84413$$

Substituting these values in equation 10 and showing intermediate values to show relative importance of terms yields

$$\int_{t_1}^{t_{24}} R\,dt=2.5[11.628(0.037940)-(1.0009)(0.034707)+6.850+6.425]$$
$$=34.202 \text{ hours cumulative aging from } t_1 \text{ to } t_{24} \qquad (11)$$

Equation 10 contains exponential integrals of the form

$$Ei(x)=\int_{-\infty}^{x}\frac{\epsilon^n}{n}\,dn \qquad (12)$$

These are evaluated by tables in references 6 and 7.

Comparison of Aging for Time-Temperature Cycles: Cases 1, 2, and 3

Summarizing the conditions of this comparison:

Initial temperature $T_0=140$ C (413 K).
Top temperature attained $T_1=180$ C (453 K)(452.905 calculated for cases 3a and 3b).
Elapsed time with heating load $L_1=t_1-t_0=$ 4 hours.
Elapsed time with cooling load $L_0=t_{24}-t_1=$ 20 hours for cases 1 and 3b; 4 hours to T_0+16 hours at T_0 for case 2.
Final temperature at end of 24 hours $=T_0=$ 140 C (413 K)(413.013 for case 3b).
Time constant $=1/\alpha=2.5$.

The cumulative relative aging from combined exponential heating and cooling cycle, case 3 is the sum of equations 8 and 11 = 23.688 + 34.202 = 57.890 hours.

$$\text{Relative aging over 24-hour cycle of Fig. 1}=\frac{57.890}{24}=2.412 \qquad (13)$$

These three cases for the conditions compared gave results as found in Table I.

To have a broader base for comparison, calculations similar to the foregoing were made with elapsed time of heating load L_1 with values from $t_1=1$, to $t_1=20$ (only to $t_1=12$ for the triangular case, since for t_1 greater than 12 the time t_1 to t_{24} would be insufficient to reach T_0 for the cooling portion). Also note that with exponential cooling the temperature at the end of the 24-hour cycle (t_{24}) was 414.36 K with a 15-hour heating cycle, and 423.09 K for 20 hours' heating (it did not return to $T_0=413$ K). The top temperature used in each case was the calculated value obtained with the use of equation 1, (and a rounded off value for cases 1 and 2). These temperature cycles are graphically shown in Fig. 6.

The results of aging calculations for various time-temperature cycles are shown in Table II and in Fig. 7. The percent errors with the use of the approximations of cases 1 and 2 compared to the theoretically correct are shown in Table III.

Application to Aging Problems

In actual aging problems, which involve a variety of parameters, the calculation of results from the theoretical equations will be rather tedious. Some progress has been made in the development of graphical methods based on these equations. It is expected that these graphical simplifications will be available in the not too distant future. Computer programming has been considered and the detailed appendixes allow this to be picked up at the proper place as indicated at the end of each appendix.

Conclusions

1. Equations have been developed giving relative insulation aging for conditions of both a temperature change at a constant rate and temperature change at an exponential rate.

2. These equations incorporate the theo-

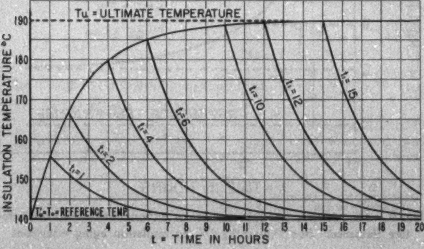


Fig. 6. Insulation temperature cycles used in comparing aging calculations based on various assumptions

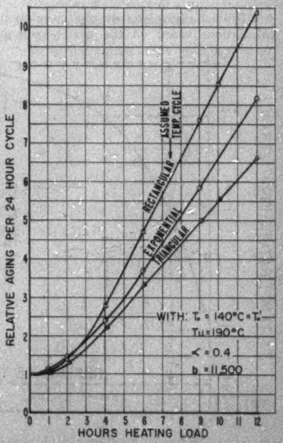


Fig. 7. Comparison of relative insulation aging based on various assumed insulation temperature cycles

Table II. Comparison of Aging for Temperature Cycles 1, 2, and 3

Hours of Heating, $t_1-t_0=t$	Degrees Kelvin: Top Temperature* Attained $=T_1$	Degrees Kelvin: Final Temperature* at 24 Hours†	Relative Aging Over 24-Hour Cycle: Case 1, Fig. 2	Relative Aging Over 24-Hour Cycle: Case 2, Fig. 3	Relative Aging Over 24-Hour Cycle: Case 3, Fig. 1
1	429.484	413.00166	1.08	1.05	1.153
2	440.533	413.00416	1.39	1.29	1.448
4	452.905	413.01337	2.78	2.17	2.412
6	458.464	413.03396	4.70	3.32	3.663
9	461.634	413.12056	7.62	5.00	5.846
10	462.084	413.18151	8.60	5.55	6.619
12	462.588	413.40811	10.39	6.63	8.180
15	462.876	414.36281	12.93		10.554
20	462.983	423.09146	17.01		14.481

* These are carried out to six significant figures to allow accurate calculations of exponential temperature cycles. There are eight significant figures for final temperatures.

† Final temperature is given as calculated for exponential temperature cycles to allow accurate calculation. In Cases 1 and 2 the final temperature used was $T_0=413$ K.

In all cases $T_0=140+273=413$ K; $1/\alpha=2.5$ hours.

Table III. Per-Cent Error in Use of Approximations

Hours of Heating, $t_1-t_0=t$	Relative Aging Over 24-Hour Cycle, From Table II: Case 1	Case 2	Case 3	Per-Cent Error Compared with Case 3: Case 1	Case 2
1	1.08	1.05	1.153	− 6.32	− 8.93
2	1.39	1.29	1.448	− 4.01	−10.9
4	2.78	2.17	2.412	+15.3	−10.0
6	4.70	3.32	3.663	+28.3	− 9.4
9	7.62	5.00	5.846	+30.4	−14.5
10	8.60	5.55	6.619	+29.9	−16.1
12	10.39	6.63	8.180	+27.0	−19.0
15	12.93		10.554	+22.5	
20	17.01		14.481	+17.5	

retically sound and generally accepted Arrhenius' chemical reaction rate concept applied to insulation deterioration as contrasted to the rough approximation of a "degree rule."

3. Typical examples show that insulation calculations based on the assumption of a triangular temperature cycle, replacing the much more accurate exponential temperature cycle, give answers of the order of 10% too small.

4. Typical examples show that insulation calculations based on the quite common assumption of a rectangular temperature cycle, replacing the much more accurate exponential temperature cycle, give answers that are too small by the order of 5% for short time loads to as much as 30% too large for longer time loads.

5. Calculations based on exponential temperature change as is generally experienced by apparatus can now be more accurately applied in investigating the effect of insulation aging in regard to apparatus time constant, loading practices, insulations with different aging rates, and insulations of different temperature classes.

Appendix I. Nomenclature

$A=T_u-T_0'$, ultimate temperature rise above T_0'; see Fig. 8
$D=T_1-T_0'$, ultimate temperature fall below T_1; see Fig. 9
$L=b/T_1$ (simplifying nomenclature)
$M=b/T_0+c$ (simplifying nomenclature)
$N=b/T_0+c+D\epsilon^{-\alpha t}$ (simplifying nomenclature)
R = relative insulation aging with unity assigned to aging at reference temperature T_0
T = insulation temperature (C+273) at any given time t, K
T_1 = insulation temperature at $t=t_1$ (final temperature for increasing temperature cycle and initial temperature for decreasing temperature cycle), K
T_0' = initial insulation temperature at $t=t_0=0$ for Fig. 8; or $t=\infty$ for Fig. 9, K
T_0 = reference insulation temperature, K. (Normally this would be rated hottest spot temperature for insulation used considering 30 C ambient)
T_u = ultimate maximum temperature of insulation for particular load for $t=\infty$; see Fig. 8, K
$a=T_0$ (used in equations to indicate that this is a reference quantity independent of time)
b = slope of relative aging curve; see references 4 and 5 for representative values)
$c=T_0'-T_0$; see Figs. 5, 8, and 9
d = differential, with adjacent variable
ϵ = base of Naperian logarithms = 2.71828
f = per-hour rate of temperature change, C
$m=b/T_0+c+D\epsilon^{-\alpha t}$ (simplifying nomenclature)
$n=(D\epsilon^{-\alpha t}m)/(T_0-c)$ (simplifying nomenclature)
$p=b/T_0+c$ (simplifying nomenclature)
$r=b/A+T_0+c$ (simplifying nomenclature)
$s=b/T_0+c+A(1-\epsilon^{-\alpha t})$ (simplifying nomenclature)
t = time, hours
t_0 = time at $t=0$
$u=b/T_0+c+ft$ (simplifying nomenclature)
$u_0=b/T_0+c$ (value of u at $t=0$)

Vol.
Pt. 1
1959

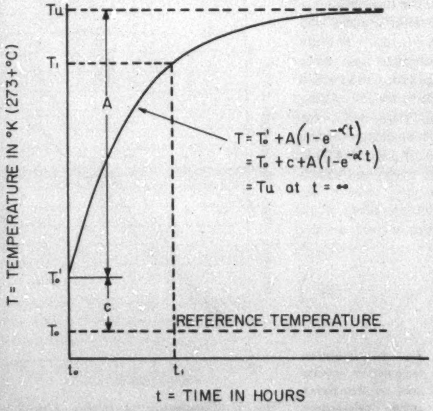


Fig. 8. Time-temperature relations for an exponential rate of temperature increase. See Appendix III for insulation aging equations

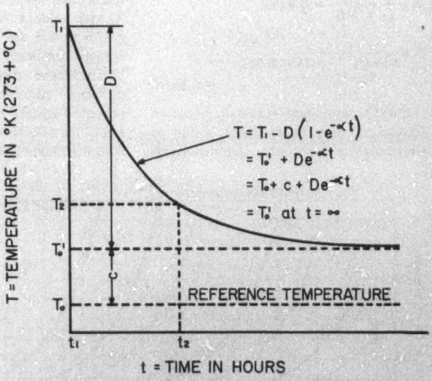


Fig. 9. Time-temperature relations for an exponential rate of temperature decrease. See Appendix IV for insulation aging equations

$x = b/T_0 + c + A(1-\epsilon^{-\alpha t})$ (simplifying nomenclature)
$y = (xA\epsilon^{-\alpha t})/(T_0+c+A)$ (simplifying nomenclature)
$1/\alpha$ = time constant of the apparatus (time to reach 63.2% of T_u)
ln = logarithm to base ϵ

Appendix II. Insulation-Aging Calculation with Constant Rate of Temperature Increase

In this evaluation insulation aging will be computed in terms of relative aging in the manner proposed by Sumner, Stein, and Lockie.[8] This method assigns a factor of unity to the insulation deterioration at some reference temperature. Aging at any other temperature is then some fraction or some multiple of the aging at this specified temperature, depending on whether the temperature is lower or higher than the specified base temperature.

Insulation aging will also be considered to be a function of the chemical reaction rate theory of Arrhenius[9] as suggested by Akahira,[10,11] Malmlow,[12] and Dakin.[3]

Combining these ideas results in the equation

$$\ln R = b\left(\frac{1}{T_0} - \frac{1}{T}\right) \tag{14}$$

Graphs of this equation are given in Fig. 4 with two representative values of b and T_0. Equation 1 may be written

$$R = \epsilon^{\left(\frac{b}{T_0} - \frac{b}{T}\right)} \tag{15}$$

$$= \epsilon^{\left(\frac{b}{T_0}\right)} \epsilon^{-\left(\frac{b}{T}\right)} \tag{16}$$

For the constant rate of temperature increase, Fig. 5 shows the relation between temperature and time. From Fig. 5

$$T_0' = T_0 + c \tag{17}$$

$$T = T_0' + ft \tag{18}$$

Then

$$T = T_0 + c + ft \tag{19}$$

Let

$$a = T_0 \text{ (since } T_0 \text{ is independent of time)} \tag{20}$$

Then

$$T = a + c + ft \tag{21}$$

Substituting equations 20 and 21 in equation 16 yields

$$R = \epsilon^{\left(\frac{b}{a}\right)} \epsilon^{-\left(\frac{b}{a+c+ft}\right)} \tag{22}$$

The cumulative relative aging from time $t=0$ to $t=t$ is $\int_0^t R\,dt$. Then

$$\int_0^t R\,dt = \epsilon^{\left(\frac{b}{a}\right)} \int_0^t \epsilon^{-\left(\frac{b}{a+c+ft}\right)} dt \tag{23}$$

Let

$$u = \frac{b}{a+c+ft} \tag{24}$$

Then

$$du = -\frac{bf}{(a+c+ft)^2}\,dt$$

$$= -\frac{b^2 f}{b(a+c+ft)^2}\,dt \left(\text{multiplying by } \frac{b}{b}\right)$$

$$= -\frac{f}{b}\,u^2 dt \text{ (from equation 24)}$$

Therefore

$$dt = -\frac{b}{f}\frac{du}{u^2} \tag{25}$$

Therefore, substituting equations 24 and 25 in equation 23 yields

$$\int_0^t R\,dt = \epsilon^{\left(\frac{b}{a}\right)} \int_0^t \epsilon^{-u}\left(-\frac{b}{f}\frac{du}{u^2}\right)$$

$$= -\frac{b}{f}\,\epsilon^{\left(\frac{b}{a}\right)} \int_0^t \frac{\epsilon^{-u}}{u^2}\,du \tag{26}$$

The right-hand integral must be integrated by parts. Let

$$\int \frac{\epsilon^{-u}}{u^2}\,du = \int w\,dv \tag{27}$$

By the well-known integration by parts transformation

$$\int w\,dv = wv - \int v\,dw \tag{28}$$

Let

$$w = \epsilon^{-u} \tag{29}$$

Then

$$dw = -\epsilon^{-u}\,du \tag{30}$$

Let

$$dv = \frac{du}{u^2} \tag{31}$$

Then

$$v = \int \frac{du}{u^2} = -\frac{1}{u} \tag{32}$$

Substituting equations 29, 30, 31, and 32 in equation 28 yields

$$\int w\,dv = -\frac{\epsilon^{-u}}{u} - \int \frac{\epsilon^{-u}}{u}\,du \tag{33}$$

Substituting equation 33 in equation 27 yields

$$\int \frac{\epsilon^{-u}}{u^2}\,du = -\frac{\epsilon^{-u}}{u} - \int \frac{\epsilon^{-u}}{u}\,du \tag{34}$$

Substituting equation 34 in equation 26 yields

$$\int_0^t R\,dt = -\frac{b}{f}\,\epsilon^{\left(\frac{b}{a}\right)} \left(-\frac{\epsilon^{-u}}{u} - \int \frac{\epsilon^{-u}}{u}\,du\right)_0^t$$

$$= +\frac{b}{f}\,\epsilon^{\left(\frac{b}{a}\right)} \left(\left.\frac{\epsilon^{-u}}{u}\right|_0^t + \int_0^t \frac{\epsilon^{-u}}{u}\,du\right) \tag{35}$$

Using equation 24 to determine boundary conditions, and noting from equation 20 that $a = T_0$, yields at $t=0$

$$u = \frac{b}{a+c} = \frac{b}{T_0+c} = u_0 \tag{36}$$

and at $t=t$

$$u = \frac{b}{a+c+ft} = \frac{b}{T_0+c+ft} \tag{37}$$

Substituting equations 20, 36, and 37 in equation 35 yields

$$\int_0^t R\,dt = \frac{b}{f}\,\epsilon^{\left(\frac{b}{T_0}\right)} \left(\left.\frac{\epsilon^{-u}}{u}\right|_{u_0}^{u} + \int_{u_0}^{u} \frac{\epsilon^{-u}}{u}\,du\right)$$

$$\int_0^t R\,dt = \frac{b}{f}\,\epsilon^{\left(\frac{b}{T_0}\right)} \left[\left(\frac{\epsilon^{-u}}{u} - \frac{\epsilon^{-u_0}}{u_0}\right) + \int_{u_0}^{u} \frac{\epsilon^{-u}}{u}\,du\right] \tag{38}$$

The right-hand integral can be evaluated by changing into the form of an exponential integral, $-Ei(-x)$, by the following transformation:

$$\int_{u_0}^{u} = \int_{u_0}^{\infty} - \int_{u}^{\infty} \tag{39}$$

Using equation 39 in equation 38 yields

$$\int_0^t R\,dt = \frac{b}{f}\,\epsilon^{\left(\frac{b}{T_0}\right)} \left(\frac{\epsilon^{-u}}{u} - \frac{\epsilon^{-u_0}}{u_0} + \int_{u_0}^{\infty} \frac{\epsilon^{-u}}{u}\,du - \int_{u}^{\infty} \frac{\epsilon^{-u}}{u}\,du\right) \tag{40}$$

Tables of these integrals are available for values of $u = 0$ to 15 in reference 6 and $u = 20$ to 50 in reference 7. However, u and u_0 are usually of the order of 20 to 30 and the reference tables are in rather large increments (also, this reference is not generally available). For values of $u > 10$ the following expression may be used with an error of less than 1.0% (less than 0.05% for $u > 20$):

$$\int_u^{\infty} \frac{\epsilon^{-u}}{u}\,du = \epsilon^{-u}\left(\frac{1}{u} - \frac{1}{u^2} + \frac{2!}{u^3} - \frac{3!}{u^4} + \ldots\right) \tag{41}$$

The difference between the true value of the integral and the sum of a finite number of the terms used is of the order of the last term used.

Substituting equation 41 in equation 40 yields

$$\int_0^t R\,dt = \frac{b}{f}\,\epsilon^{\left(\frac{b}{T_0}\right)} \left[\frac{\epsilon^{-u}}{u} - \frac{\epsilon^{-u_0}}{u_0} + \epsilon^{-u_0}\left(\frac{1}{u_0} - \frac{1}{u_0^2} + \frac{2!}{u_0^3} - \frac{3!}{u_0^4}\right) - \epsilon^{-u}\left(\frac{1}{u} - \frac{1}{u^2} + \frac{2!}{u^3} - \frac{3!}{u^4}\right)\right] \tag{42}$$

This simplifies to

$$\int_0^t R\,dt = \frac{b}{f}\,\epsilon^{\left(\frac{b}{T_0}\right)} \left[\epsilon^{-u}\left(\frac{1}{u^2} - \frac{2}{u^3} + \frac{6}{u^4}\right) - \epsilon^{-u_0}\left(\frac{1}{u_0^2} - \frac{2}{u_0^3} + \frac{6}{u_0^4}\right)\right]$$

$$= \frac{b}{f}\left[\epsilon^{\left(\frac{b}{T_0} - u\right)}\left(\frac{1}{u^2} - \frac{2}{u^3} + \frac{6}{u^4}\right) - \epsilon^{\left(\frac{b}{T_0} - u_0\right)}\left(\frac{1}{u_0^2} - \frac{2}{u_0^3} + \frac{6}{u_0^4}\right)\right] \tag{43}$$

where, from equations 36 and 37,

$$u_0 = \frac{b}{T_0+c} \text{ and } u = \frac{b}{T+c+ft}$$

A frequent case is where $c=0$, meaning that $T_0'=T_0$, in which case note that equation 43 becomes

$$\int_0^t R\,dt=\frac{b}{f}\left[\epsilon^{u_0-u}\left(\frac{1}{u^2}-\frac{2}{u^3}+\frac{6}{u^4}\right)-\epsilon^{u_0-u_0}\left(\frac{1}{u_0^2}-\frac{2}{u_0^3}+\frac{6}{u_0^4}\right)\right]$$

$$=\frac{b}{f}\left[\epsilon^{u_0-u}\left(\frac{1}{u^2}-\frac{2}{u^3}+\frac{6}{u^4}\right)-\left(\frac{1}{u_0^2}-\frac{2}{u_0^3}+\frac{6}{u_0^4}\right)\right] \quad (44)$$

where $u_0=(b/T_0)$; $u=b/(T_0+ft)$; and $c=0$.

For computer programming use equation 40 with infinite series, equation 41. (The number of terms used must not be greater than u.)

Appendix III. Insulation Aging Calculation with Exponential Temperature Increase

Assume an exponential temperature increase in accordance with Fig. 8 with the relation

$$T=T_0'+A(1-\epsilon^{-\alpha t}) \quad (45)$$

If $T_0'=T_0+c$ (see Fig. 8) equation 45 becomes

$$T=T_0+c+A(1-\epsilon^{-\alpha t}) \quad (46)$$

Using relative aging relations as developed in Appendix II, and starting with equation 16 in Appendix II, yields

$$R=\epsilon^{\left(\frac{b}{T_0}\right)}\epsilon^{-\left(\frac{b}{T}\right)} \quad (47)$$

Substituting equation 46 in equation 47 yields

$$R=\epsilon^{\left(\frac{b}{T_0}\right)}\epsilon^{-\left(\frac{b}{T_0+c+A(1-\epsilon^{-\alpha t})}\right)} \quad (48)$$

The cumulative relative aging from time $t=0$ to $t=t$ is $\int_0^t R\,dt$.

$$\int_0^t R\,dt=\epsilon^{\frac{b}{T_0}}\int_0^t \epsilon^{-\frac{b}{T_0+c-A(1-\epsilon^{-\alpha t})}}\,dt \quad (49)$$

Let

$$x=\frac{b}{T_0+c+A(1-\epsilon^{-\alpha t})} \quad (50)$$

Then

$$dx=\frac{-b\alpha A\epsilon^{-\alpha t}}{[T_0+c+A(1-\epsilon^{-\alpha t})]^2}\,dt \quad (51)$$

$$=\frac{-b^2\alpha A\epsilon^{-\alpha t}\,dt}{b[T_0+c+A(1-\epsilon^{-\alpha t})]^2}\,dt$$

(multiplying by b/b).

$$dx=\frac{-\alpha A\epsilon^{-\alpha t}}{b}x^2\,dt \quad (51A)$$

(from 50 and 51). But $T_0+c+A-A\epsilon^{-\alpha t}=b/x$ (from equation 50). Therefore

$$\epsilon^{-\alpha t}=-\frac{\frac{b}{x}-T_0-c}{A}+1 \quad (52)$$

Substituting equation 52 in equation 51(A) yields

$$dx=-\frac{x^2A\alpha}{b}\left[-\frac{\frac{b}{x}-T_0-c}{A}+1\right]dt$$

$$=x\alpha\left[1+x\left(-\frac{A+T_0+c}{b}\right)\right]dt \quad (53)$$

Therefore

$$dt=\frac{dx}{x\alpha\left[1+x\left(-\frac{A+T_0+c}{b}\right)\right]} \quad (54)$$

Let

$$g=-\left(\frac{A+T_0+c}{b}\right) \quad (55)$$

Then

$$dt=\frac{dx}{x\alpha(1+gx)} \quad (56)$$

Substituting equations 50 and 56 in equation 49

$$\int_0^t R\,dt=\frac{\epsilon^{\frac{b}{T_0}}}{\alpha}\int_{x_0}^{x}\frac{\epsilon^{-x}\,dx}{x(1+gx)} \quad (57)$$

Let

$$h=\frac{\epsilon^{\frac{b}{T_0}}}{\alpha} \quad (58)$$

Let

$$\frac{1}{x(1+gx)}=\frac{B}{x}+\frac{E}{1+gx}$$

from which $B=1$ and $E=-g$, resulting in

$$\frac{1}{x(1+gx)}=\frac{1}{x}-\frac{g}{1+gx} \quad (59)$$

Substituting equations 58 and 59 in equation 57

$$\int_0^t R\,dt=h\int_{x_0}^{x}\epsilon^{-x}\left[\frac{1}{x}-\frac{g}{1+gx}\right]dx$$

$$=h\int_{x_0}^{x}\frac{\epsilon^{-x}}{x}\,dx-h\int_{x_0}^{x}\frac{g\epsilon^{-x}}{1+gx}\,dx \quad (60)$$

Let

$$F=-h\int_{x_0}^{x}\frac{g\epsilon^{-x}}{1+gx}\,dx \quad (61)$$

Also let

$$z=1+gx \quad (62)$$

Then

$$dz=g\,dx$$

$$dx=\frac{dz}{g}$$

$$x=\frac{z-1}{g}$$

from equation 62.

Make these substitutions in equation 61. Then

$$F=-h\int_{z_0}^{z}\frac{g\epsilon^{-\frac{z-1}{g}}}{z}\frac{dz}{g} \quad (63)$$

$$=-h\epsilon^{\frac{1}{g}}\int_{z_0}^{z}\frac{\epsilon^{-\frac{z}{g}}}{z}\,dz \quad (64)$$

Let

$$y=\frac{z}{g} \quad (65)$$

Then

$$dy=\frac{dz}{g}$$

$$dz=g\,dy$$

from equation 65.

$$z=gy$$

Substituting equation 65 in equation 64 yields

$$F=-h\epsilon^{\frac{1}{g}}\int_{y_0}^{y}\frac{\epsilon^{-y}}{gy}\,g\,dy$$

$$=-h\epsilon^{\frac{1}{g}}\int_{y_0}^{y}\frac{\epsilon^{-y}}{y}\,dy \quad (66)$$

Substituting equation 66 in equation 61, then equation 61 in equation 60 yields

$$\int_0^t R\,dt=h\int_{x_0}^{x}\frac{\epsilon^{-x}}{x}\,dx-h\epsilon^{\frac{1}{g}}\int_{y_0}^{y}\frac{\epsilon^{-y}}{y}\,dy \quad (67)$$

Evaluating limits on the first integral of equation 67, when $t=0$, yields

$$x=\frac{b}{T_0+c}=x_0 \quad (68)$$

from equation 50. When $t=t$

$$x=\frac{b}{T_0+c+A(1-\epsilon^{-\alpha t})} \quad (69)$$

from equation 50.

Evaluating limits on the second integral of equation 67 yields

$$y=\frac{z}{g}$$

from equation 65;

$$=\frac{1+gx}{g}$$

from equation 62;

$$=\frac{1}{g}+x$$

$$=-\frac{b}{A+T_0+c}+\frac{b}{T_0+c+A(1-\epsilon^{-\alpha t})}$$

from equations 50 and 55.

When $t=0$

$$y=-\frac{b}{A+T_0+c}+\frac{b}{T_0+c}=y_0 \quad (70)$$

When $t=t$

$$y=-\frac{b}{A+T_0+c}+\frac{b}{T_0+c+A(1-\epsilon^{-\alpha t})} \quad (71)$$

Substitute limits given, equations 68 through 71, in equation 67, and note that

$$h=\frac{\epsilon^{\frac{b}{T_0}}}{\alpha}$$

from equation 58, and that

$$\epsilon^{\frac{1}{\theta}}=\epsilon^{-\frac{b}{A+T_0+c}}$$

from equation 55, yielding

$$\int_0^t Rdt=\frac{\epsilon^{\frac{b}{T_0}}}{\alpha}\left[\int_{\frac{b}{T_0+c}}^{\frac{b}{T_0+c+A(1-\epsilon^{-\alpha t})}} \times \frac{\epsilon^{-x}}{x}\,dx-\epsilon^{-\frac{b}{A+c+T_0}}\times \int_{-\frac{b}{A+c+T_0}+\frac{b}{T_0+c}}^{-\frac{b}{A+c+T_0}+\frac{b}{T_0+c+A(1-\epsilon^{-\alpha t})}} \frac{\epsilon^{-y}}{y}dy\right] \quad (72)$$

To simplify nomenclature, let

$$p=\frac{b}{T_0+c} \quad (73)$$

$$r=\frac{b}{A+T_0+c} \quad (74)$$

$$s=\frac{b}{T_0+c+A(1-\epsilon^{-\alpha t})} \quad (75)$$

Then equation 72 becomes

$$\int_0^t Rdt=\frac{\epsilon^{\frac{b}{T_0}}}{\alpha}\left[\int_p^s\frac{\epsilon^{-x}}{x}\,dx-\epsilon^{-r}\times \int_{-r+p}^{-r+s}\frac{\epsilon^{-y}}{y}dy\right] \quad (76)$$

The integrals in equation 76 may be put in the form of $[-Ei(-x)]$ functions (exponential integrals)[8] by using the following relations:

$$\int_{x_0}^x\frac{\epsilon^{-x}}{x}\,dx=\underset{a\to\infty}{\text{limit}}\left[\int_{x_0}^a\frac{\epsilon^{-x}}{x}\,dx+\int_a^x\frac{\epsilon^{-x}}{x}\,dx\right]$$

$$=\int_{x_0}^\infty\frac{\epsilon^{-x}}{x}\,dx+\int_\infty^x\frac{\epsilon^{-x}}{x}\,dx$$

$$\int_{x_0}^x\frac{\epsilon^{-x}}{x}\,dx=\int_{x_0}^\infty\frac{\epsilon^{-x}}{x}\,dx-\int_x^\infty\frac{\epsilon^{-x}}{x}\,dx \quad (77)$$

Using equation 77 in equation 76 yields

$$\int_0^t Rdt=\frac{\epsilon^{\frac{b}{T_0}}}{\alpha}\left[\int_p^\infty\frac{\epsilon^{-x}}{x}\,dx-\int_s^\infty\frac{\epsilon^{-x}}{x}\,dx-\epsilon^{-r}\left(\int_{p-r}^\infty\frac{\epsilon^{-y}}{y}\,dy-\int_{s-r}^\infty\frac{\epsilon^{-y}}{y}\,dy\right)\right] \quad (78)$$

Equation 78 is in the form of $[-Ei(-x)]$ functions and can thus be evaluated by reference to tables of exponential integrals.[6,7]

Since the values of p and s in equation 78 are usually greater than 20, and since tables of the $[-Ei(-x)]$ functions are in large increments, or in some cases not available, use is made of equation 41 to replace the first two integrals of equation 78. Equation 78 then becomes

$$\int_0^t Rdt=\frac{\epsilon^{\frac{b}{T_0}}}{\alpha}\left[\epsilon^{-p}\left(\frac{1}{p}-\frac{1}{p^2}+\frac{2}{p^3}-\frac{6}{p^4}\right)-\epsilon^{-s}\left(\frac{1}{s}-\frac{1}{s^2}+\frac{2}{s^3}-\frac{6}{s^4}\right)-\epsilon^{-r}\left(\int_{p-r}^\infty\frac{\epsilon^{-y}}{y}\,dy-\int_{s-r}^\infty\frac{\epsilon^{-y}}{y}\,dy\right)\right] \quad (79)$$

For the case of $T_0'=T$ $(c=0)$, $b/T_0=p$. Hence

$$\int_0^t Rdt=\frac{\epsilon^p}{\alpha}\left[\epsilon^{-p}\left(\frac{1}{p}-\frac{1}{p^2}+\frac{2}{p^3}-\frac{6}{p^4}\right)-\epsilon^{-s}\left(\frac{1}{s}-\frac{1}{s^2}+\frac{2}{s^3}-\frac{6}{s^4}\right)-\epsilon^{-r}\left(\int_{p-r}^\infty\frac{\epsilon^{-y}}{y}\,dy-\int_{s-r}^\infty\frac{\epsilon^{-y}}{y}\,dy\right)\right]$$

$$\int_0^t Rdt=\frac{1}{\alpha}\left[1\left(\frac{1}{p}-\frac{1}{p^2}+\frac{2}{p^3}-\frac{6}{p^4}\right)-\epsilon^{p-s}\left(\frac{1}{s}-\frac{1}{s^2}+\frac{2}{s^3}-\frac{6}{s^4}\right)-\epsilon^{p-r}\left(\int_{p-r}^\infty\frac{\epsilon^{-y}}{y}dy-\int_{s-r}^\infty\frac{\epsilon^{-y}}{y}\,dy\right)\right] \quad (80)$$

Caution: Use equation 80 only if $T_0'=T_0$; otherwise use equation 79.

Note that $\int_z^\infty\frac{\epsilon^{-y}}{y}\,dy$ in equations 79 and 80 are in the form of exponential integrals which may be evaluated by tables.[6,7] For computer programming use equation 78 with infinite series 41. (The number of terms used must not be greater than u.)

Appendix IV. Insulation-Aging Calculation with Exponential Temperature Decrease

Assume an exponential temperature decrease in accordance with Fig. 9, with the relation

$$T=T_1-D(1-\epsilon^{-\alpha t})$$
$$=T_0'+De^{-\alpha t}$$
$$=T_0+c+De^{-\alpha t} \quad (81)$$

Using relative aging relations as developed in Appendix II, and starting with equation 16 yields

$$R=\epsilon^{\frac{b}{T_0}}\ \epsilon^{-\frac{b}{T}} \quad (82)$$

Substituting equation 81 in equation 82 yields

$$R=\epsilon^{\frac{b}{T_0}}\ \epsilon^{-\frac{b}{T_0+c+De^{-\alpha t}}} \quad (83)$$

The cumulative relative aging from time $t=0$ to $t=t$ is $\int_0^t Rdt$.

$$\int_0^t Rdt=\epsilon^{\frac{b}{T_0}}\int_0^t\epsilon^{-\frac{b}{T_0+c+De^{-\alpha t}}}dt \quad (84)$$

Let

$$m=\frac{b}{T_0+c+D\epsilon^{-\alpha t}} \quad (85)$$

Then

$$\epsilon^{-\alpha t}=\frac{b}{mD}-\frac{T_0+c}{D} \quad (86)$$

Differentiating,

$$-\alpha\epsilon^{-\alpha t}\,dt=-\frac{bD}{(mD)^2}\,dm$$

$$dt=\frac{b}{m^2D\alpha\epsilon^{-\alpha t}}\,dm \quad (87)$$

Substituting equation 86 in 87 yields

$$dt=\frac{b\,dm}{m^2D\alpha\left(\frac{b}{mD}-\frac{T_0+c}{D}\right)}$$

$$=\frac{dm}{\alpha m\left[1-m\frac{(T_0+c)}{b}\right]} \quad (88)$$

Let

$$k=\frac{T_0+c}{b} \quad (89)$$

Substituting equation 89 in 88 yields

$$dt=-\frac{dm}{\alpha m(1-km)} \quad (90)$$

Substituting equations 85 and 90 in equation 84 yields

$$\int_0^t Rdt=\epsilon^{\frac{b}{T_0}}\int_{m_0}^m\epsilon^{-m}\frac{dm}{\alpha m(1-km)}$$

$$=\frac{\epsilon^{\frac{b}{T_0}}}{\alpha}\int_{m_0}^m\epsilon^{-m}\frac{dm}{m(1-km)} \quad (91)$$

Let

$$\frac{1}{m(1-km)}=\frac{G}{m}+\frac{H}{1-km}$$

from which $G=1$ and $H=k$, resulting in

$$\frac{1}{m(1-km)}=\frac{1}{m}+\frac{k}{1-km} \quad (92)$$

Substituting equation 92 in 91 yields

$$\int_0^t Rdt=\frac{\epsilon^{\frac{b}{T_0}}}{\alpha}\left[\int_{m_0}^m\epsilon^{-m}\,dm\left(\frac{1}{m}+\frac{k}{1-km}\right)\right]$$

$$=\frac{\epsilon^{\frac{b}{T_0}}}{\alpha}\left[\int_{m_0}^m\epsilon^{-m}\frac{dm}{m}+\int_{m_0}^m\frac{k\epsilon^{-m}dm}{1-km}\right] \quad (93)$$

Let

$$J=\int_{m_0}^m\frac{k\epsilon^{-m}dm}{1-km} \quad (94)$$

Let

$$kn=1-km$$

$$n=\frac{1}{k}-m \quad (95)$$

$$m=\frac{1}{k}-n \qquad (96)$$

$$dm=-dn \qquad (97)$$

Substituting equations 95 through 97 in equation 94

$$J=\int_{n_0}^{n}\frac{k\epsilon^{-\left(\frac{1}{k}-n\right)}(-dn)}{kn}$$

$$=-\epsilon^{-\frac{1}{k}}\int_{n_0}^{n}\frac{\epsilon^{n}}{n}dn \qquad (98)$$

Substituting equation 98 in equations 93 and 94

$$\int_0^t Rdt=\frac{\epsilon^{-\frac{b}{T_0}}}{\alpha}\left[\int_{m_0}^{m}\frac{\epsilon^{-m}}{m}dm-\epsilon^{-\frac{1}{k}}\times \int_{n_0}^{n}\frac{\epsilon^{n}}{n}dn\right] \qquad (99)$$

Evaluating limits on the first integral of equation 99 from equation 85, when $t=0$,

$$m=\frac{b}{T_0+c+D}=\frac{b}{T_1}=m_0 \qquad (100)$$

and when $t=t$

$$m=\frac{b}{T_0+c+D\epsilon^{-\alpha t}} \qquad (101)$$

Evaluating limits on the second integral of equation 99 from equation 95

$$n=\frac{1}{k}-m$$

Substituting in this equations 85 and 89, when $t=t$,

$$n=\frac{b}{T_0+c}-\frac{b}{T_0+c+D\epsilon^{-\alpha t}} \qquad (102)$$

and when $t=0$,

$$n_0=\frac{b}{T_0+c}-\frac{b}{T_1} \qquad (103)$$

where $T_1=T_0+c+D$

Substituting equations 89 and 100 through 103 in equation 99 yields

$$\int_0^t Rdt=\frac{\epsilon^{\frac{b}{T}}}{\alpha}\left[\int_{\frac{b}{T_1}}^{\frac{b}{T_0+c+D\epsilon^{-\alpha t}}}\frac{\epsilon^{-m}}{m}dm-\epsilon^{-\frac{b}{T_0+c}}\int_{\frac{b}{T_0+c}-\frac{b}{T_1}}^{\frac{b}{T_0+c}-\frac{b}{T_0+c+D\epsilon^{-\alpha t}}}\frac{\epsilon^{n}}{n}dn\right] \qquad (104)$$

Let

$$L=\frac{b}{T_1} \qquad (105)$$

$$M=\frac{b}{T_0+c} \qquad (106)$$

$$N=\frac{b}{T_0+c+D\epsilon^{-\alpha t}} \qquad (107)$$

Substituting equations 105 through 107 in equation 104 yields

$$\int_0^t Rdt=\frac{\epsilon^{\frac{b}{T_0}}}{\alpha}\left[\int_L^N\frac{\epsilon^{-m}}{m}dm-\epsilon^{-M}\int_{M-L}^{M-N}\frac{\epsilon^{n}}{n}dn\right] \qquad (108)$$

The first integral in equation 108 may be put in the form of $(-Ei)(-X)$ functions (exponential integrals) by the use of the relation in equation 77.

Also the second integral in equation 108 may be put in the form of $(+Ei)(+X)$ functions (exponential integrals) by use of the following relation:

$$\int_{x_0}^{x}\frac{\epsilon^{x}}{x}dx=\lim_{a\to-\infty}\left[\int_{x_0}^{a}\frac{\epsilon^{x}}{x}dx+\int_a^x\frac{\epsilon^{x}}{x}dx\right]$$

$$=\int_{x_0}^{-\infty}\frac{\epsilon^{x}}{x}dx+\int_{-\infty}^{x}\frac{\epsilon^{x}}{x}dx$$

$$=-\int_{-\infty}^{x_0}\frac{\epsilon^{x}}{x}dx+\int_{-\infty}^{x}\frac{\epsilon^{x}}{x}dx \qquad (109)$$

Substituting equations 77 and 109 in equation 108 yields

$$\int_0^t Rdt=\frac{\epsilon^{\frac{b}{T_0}}}{\alpha}\left[\int_L^\infty\frac{\epsilon^{-m}}{m}dm-\int_N^\infty\frac{\epsilon^{-m}}{m}dm-\epsilon^{-M}\left(-\int_{-\infty}^{M-L}\frac{\epsilon^{n}}{n}dn+\int_{-\infty}^{M-N}\frac{\epsilon^{n}}{n}dn\right)\right] \qquad (110)$$

If L and N are greater than 10, equation 41 can be used to replace the first two integrals in equation 110 with an error of less than 1.0%. Equation 110 becomes

$$\int_0^t Rdt=\frac{\epsilon^{\frac{b}{T_0}}}{\alpha}\left[e^{-L}\left(\frac{1}{L}-\frac{1}{L^2}+\frac{2}{L^3}-\frac{6}{L^4}\right)-\epsilon^{-N}\left(\frac{1}{N}-\frac{1}{N^2}+\frac{2}{N^3}-\frac{6}{N^4}\right)-\epsilon^{-M}\left(-\int_{-\infty}^{M-L}\frac{\epsilon^{n}}{n}dn+\int_{-\infty}^{M-N}\frac{\epsilon^{n}}{n}dn\right)\right] \qquad (111)$$

For the case of $T_0'=T_0$ (i.e., $c=0$), $b/T_0=M$. Hence

$$\int_0^t Rdt=\frac{\epsilon^{M}}{\alpha}\left[\epsilon^{-L}\left(\frac{1}{L}-\frac{1}{L^2}+\frac{2}{L^3}-\frac{6}{L^4}\right)-\epsilon^{-N}\left(\frac{1}{N}-\frac{1}{N^2}+\frac{2}{N^3}-\frac{6}{N^4}\right)-\epsilon^{-M}\left\{-\int_{-\infty}^{M-L}\frac{\epsilon^{n}}{n}dn+\int_{-\infty}^{M-N}\frac{\epsilon^{n}}{n}dn\right\}\right]$$

$$=\frac{1}{\alpha}\left[\epsilon^{M-L}\left(\frac{1}{L}-\frac{1}{L^2}+\frac{2}{L^3}-\frac{6}{L^4}\right)-e^{M-N}\left(\frac{1}{N}-\frac{1}{N^2}+\frac{2}{N^3}-\frac{6}{N^4}\right)+\int_{-\infty}^{M-L}\frac{\epsilon^{n}}{n}dn-\int_{-\infty}^{M-N}\frac{\epsilon^{n}}{n}dn\right] \qquad (112)$$

Caution: Use equation 112 only if $T_0'=T$ (when $c=0$); otherwise use equation 111. Note that

$$\int_{-\infty}^{x}\frac{\epsilon^{n}}{n}dn$$

in equations 111 and 112, is an exponential integral which can be evaluated from tables.[6,7]

For computer programming use equation 110 with infinite series, equation 41, for the $(-Ei)(-X)$ functions and a series like that of equation 41, except that all signs are made positive for the $(+Ei)(+X)$ functions. (The number of terms used must not be greater than u.)

References

1. THE THERMAL TIME CONSTANTS OF DYNAMO-ELECTRIC MACHINES, **A. E. Kennelly.** *AIEE Transactions*, vol. 44, Feb. 1925, pp. 137–54.

2. PREDETERMINATION OF SELF-COOLED OIL-IMMERSED TRANSFORMER TEMPERATURES BEFORE CONDITIONS ARE CONSTANT, **W. H. Cooney.** *Ibid.*, May 1925, pp. 611–18.

3. ELECTRICAL INSULATION DETERIORATION TREATED AS A CHEMICAL RATE PHENOMENON, **Thomas W. Dakin.** *Ibid.*, vol. 67, pt. I, 1948, pp. 113–22.

4. LOADING OF SEALED DRY-TYPE TRANSFORMERS, **L. C. Whitman.** *Ibid.*, vol. 76, pt. III (*Power Apparatus and Systems*), June 1957, pp. 214–23.

5. LOADING OF VENTILATED DRY-TYPE TRANSFORMERS, **L. C. Whitman.** *Ibid.*, Dec. 1957, pp. 1077–84.

6. TABLES OF SINE, COSINE AND EXPONENTIAL INTEGRALS, VOL. I AND II, **A. N. Lowan.** *Report of Official Project 765-97-3-10*, Federal Works Agency, Works Project Administration for the City of New York. Sponsored by National Bureau of Standards, Washington, D. C., 1940.

7. TABLES OF EXPONENTIALS AND EXPONENTIAL INTEGRALS, "Scientific Papers of the Institute of Physical and Chemical Research," **Takeo Akahira.** Komagome Hongo, Tokyo, Japan, June 1929, pp. 181–215, Table 3.

8. LIFE EXPECTANCY OF OIL-IMMERSED INSULATING STRUCTURES, **W. A. Sumner, G. M. Stein, A. M. Lockie.** *AIEE Transactions*, vol. 72, pt. III, Oct. 1953, pp. 924–30.

9 OUTLINES OF PHYSICAL CHEMISTRY (book), **Farrington Daniels.** John Wiley & Sons, Inc., New York, N. Y., 1948, p. 377.

10. EFFECT OF TEMPERATURE ON THE DETERIORATION OF FIBROUS INSULATING MATERIALS, **T. Kujira, T. Akahira.** "Scientific Papers of the Institute of Physical and Chemical Research," vol. 2, 1925, pp. 223–52.

11. RELATIONS BETWEEN CARBON, HYDROGEN, AND OXYGEN CONTENTS, IN COTTON CELLULOSE UNDER THERMAL DECOMPOSITION, AND ITS WEIGHT LOSS, **Takeo Akahira.** *Ibid.*, vol. 9, 1928, pp. 165–80.

12. THERMAL AGING PROPERTIES OF CELLULOSE INSULATION MATERIALS, **G. Malmlow.** *Transactions*, Royal Institute of Technology, Stockholm, Sweden, Nov. 19, 1948.

◆

Discussion

F. H. Buller (Anaconda Wire and Cable Company, Hastings-on-Hudson, N. Y.): The authors do the industry a service in contribution to the theory of insulation aging. While their fundamental idea appears to be sound, it would seem that the usefulness of their analysis could be extended

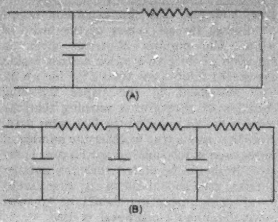


Fig. 10. Types of thermal circuits

A—For simple apparatus; see equation 45
B—For more complicated apparatus

if equation 45 were modified to represent more exactly the types of apparatus encountered in practice. This equation as it stands applies strictly speaking to the type of thermal circuit shown in Fig. 10(A).

The corresponding apparatus could be a bare conductor suspended in air, if the surface thermal resistivity of the conductor or its reciprocal (the coefficient of heat transfer) were constant over the portion of the transient under consideration.

For more complicated apparatus a more elaborate thermal circuit is needed, e.g., see Fig. 10(B). The corresponding equation is

$$T-T_0=A(1-\epsilon^{-\alpha t})+B(1-\epsilon^{-\beta t})+C(1-\epsilon^{-\gamma t})$$

The equation requires one term for each thermal capacitance in the "ladder network" as shown.

This type of circuit has been used successfully to represent an insulated cable in air, with a constant surface thermal resistivity.[1]

A generator, a motor, or a transformer can also be represented by a ladder network similar to that of Fig. 10(B), although it is possible that more meshes might be required; moreover, the approximation is not quite as good unless one or more of the meshes of the ladder network are replaced by suitable subsidiary networks. It is apparent that the corresponding equation can become quite complicated and may even require a computer for its solution.

A cable with very heavy insulation can be treated thermally as a transmission line with suitably distributed constants. In this case the foregoing equation would involve an infinite number of terms of which only the first three or four are usually of practical significance.[2]

The first term of the series is the principal term and it persists for the longest period of time; the other terms serve to modify the transient during its earlier stages. This is not necessarily true in the case of the finite series which represents a generator, a motor, or a transformer; here the various terms may or may not be of the same order of duration depending on circumstances.

Approximations of this sort could possibly be incorporated in the authors' analysis, and since in practically all of this thermal work the superposition principle holds, it might be possible to prepare term-by-term solutions.

REFERENCES

1. AN IMPROVED APPROXIMATE TECHNIQUE FOR CALCULATING CABLE TEMPERATURE TRANSIENTS, F. C. Van Wormer. *AIEE Transactions*, vol. 74, pt. III (*Power Apparatus and Systems*), Apr. 1955, pp. 277–81.

2. THE THEORY OF OIL-FILLED CABLE, G. B Shankin, F. H. Buller. *General Electric Review*, Schenectady, N. Y., vol. 34, pt. II, Sept. 1931, p. 523.

Andrew A. Halacsy (Federal Pacific Electric Company, Newark, N. J.): In connection with insulation aging we know rather more about the load than about the temperature in a given period. In practical cases, the authors' method requires the knowledge of the heating and cooling curves of the particular device at the particular load. Only by such curves can the aging be properly calculated. This is true especially because both the knowledge of the temperatures actually reached during a temperature cycle and the knowledge of T_u=ultimate maximum temperature of the insulation for the particular load are required for accurate calculation. This T_u temperature is not reached during the cycling in question nor even in a test with a heavy overload applicable, of course, only for a very short time.

This situation is unfortunate but it seems inevitable because the authors give ample proof that other simplifying methods result in unacceptable error. It must be added that, due to proximity effects in the rather complex heat transmission system of an electric apparatus, the upper part of the heating curve hardly follows an exponential curve. Therefore I suggest using a value for T_u projected from readings not higher than T_1, even if T_u can be measured by test.

An electric apparatus usually has a most dangerous spot in its insulation system, as far as aging is concerned. It is important to know which part of the insulation ages fastest under certain load conditions, because the life of the device in which we are ultimately interested is determined by the life of this particular insulation. This is the very foundation of the concept of the functional evaluation of insulation, well accepted by now. It follows that, if the authors' calculation of insulation aging is carried out for a particular material at the most dangerous or "most aging" spot, and if a load-temperature curve with a projected T_u for that material in that device under that particular load is used in the calculation, the result is an indication of the aging of the whole device. This means that in carrying through calculations of aging it is important to know not so much the evasive hot spot, but rather the most aging spot of the apparatus which might differ from the absolute hot spot, especially if insulating materials with different characteristics are used. This is sanctioned, and rightly, by the recent revision of insulation standards.

The authors are to be congratulated on their mathematics. Unfortunately, they repeatedly hit an evasive exponential integral starting in equations 33 and 67. It is recommended that these should be evaluated by reference to published tables of such integrals. Such a procedure is cumbersome, especially if the proper tables are not available, and I recommend verification, and possibly the use of equation 307 from Burlington's *Handbook*,[1] which reads as follows:

$$\frac{\epsilon^{ax}}{x}\,ax=\log x+ax+\frac{(ax)^2}{2.2!}+\frac{(ax)^3}{3.3!}+\ldots$$

In case of $a=-1$, as in this paper,

$$\frac{\epsilon^{-x}}{x}\,ax=\log x-x+\frac{x^2}{2.2!}-\frac{x^2}{3.3!}+\ldots$$

This is similar to the expansions used by the authors in other places.

Finally, I welcome the special nomenclature in the form of an appendix, and I would strongly recommend this practice in future papers.

REFERENCE

1. HANDBOOK OF MATHEMATICAL TABLES AND FORMULAS, Burlington. Handbook Publishers, Inc., Sandusky, Ohio, 1957.

L. Meyerhoff (General Cable Corporation, Bayonne, N. J.): Equation 41

$$\int_u^\infty \frac{\epsilon^{-u}}{u}\,du=\epsilon^{-u}\left(\frac{1}{u}-\frac{1}{u^2}+\frac{2!}{u^3}-\frac{3!}{u^4}+\ldots\right)$$

is stated to be a close approximation to the true value of the integral for values of u greater than 10.

The equation as shown is an infinite series which converges for a finite number of terms and then diverges. Thus, if we designate the number of terms by n, the series converges up to $n=u$ but diverges for all values of $n>u$. In other words, if the summation is carried to $n=\infty$ the sum is ∞.

It is perhaps true that if the summation is carried out for only a very limited and prescribed number of terms, such as $n=3$ or $n=u/2$, the approximation is close enough for practical purposes. In equations 42 and 43 only four terms of the series are used and it is perhaps intended never to use more than four terms. However, the last sentence of Appendix III leaves this in doubt. This says: "For computer programming use equation 78 with infinite series 41." It appears important for the authors to clarify the limitations of use of the approximate equation.

L. C. Whitman and **W. W. Whitman:** Mr. Buller has correctly pointed out that equation 45 is an approximation, as was also indicated in references 1 and 2 of the paper. Mr. Buller's additional terms give the added refinement to agree more exactly with a particular thermal circuit. At present, these additional refinements may not be justified in transformer aging calculations. An exception to this may be the case where the load on a liquid-filled unit is suddenly decreased. The thermal gradient from winding to liquid will then fall rapidly until the winding temperature approaches the liquid temperature. The time constant for this portion of the cooling cycle will thus be much different from the remainder of the cycle and could be advantageously expressed as a separate exponential. In the case of a cable with heavy insulation, additional terms are undoubtedly of practical significance.

Inclusion of these additional exponentials in the aging calculations would mean applying the same final equations such as 79 and 111 with the values of the various exponentials substituted and the results summed.

Dr. Halacsy is quite correct in that the hottest spot of equipment may not be the most dangerous spot because it may not be the spot subjected to the highest voltage or mechanical stresses, and because with the use of composite insulations of different temperature endurance, some other lower temperature spot may develop weakness before the actual hottest spot occurs. Thus we must critically examine the particular structure and determine as well as we can the most dangerously aged spot and apply our analysis to this material and region. Mr. Buller has pointed out the apparatus may not strictly follow an exponential curve. Thus, we may find, as Dr. Halacsy suggests, that altering the actual T_u (or even the T_1) used in the equations will give a better approximation of the temperature cycle, particularly if only one exponential is used.

The expansion given by Dr. Halacsy for the evaluation of the exponential integral has the disadvantage that it converges quite slowly for values of u greater than 5. Of course if a computer is used this presents no difficulties. As a further alternative we would suggest the reference at 1 of this closure as an available and inexpensive source of published tables of the exponential integral.

Mr. Meyerhoff has brought out a most important point. Several references give equation 41 as a proper expansion. It is true, however, that this expansion begins to diverge for all values of the order of $n \geq u$. The number of terms given in equation 41 do give a value correct within the stated limits (less than 1%) for $u>10$. The statement in the paper following equation 41 does give a warning that the error is of the order of the last term used. Note that this is true whether the number of terms used is too small or too large. However, it should be stated that the number of terms used should not in any case exceed $n=u$. This restriction should be imposed especially when a computer is used.

Reference

1. Mathematical Tables (book), H. B. Dwight. Dover Publications, Inc., New York, N. Y., 1958.

A Theoretical Analysis of the Effects of an Electric Field on the Charging of Fine Particles

ARTHUR T. MURPHY
ASSOCIATE MEMBER AIEE

FELIX T. ADLER
NONMEMBER AIEE

GAYLORD W. PENNEY
FELLOW AIEE

Synopsis: Recent measurements of charge imparted to fine particles by unipolar ions in a strong electric field have shown that this charge is appreciably higher than that predicted by previous theoretical analyses. The charge acquired by fine particles is due primarily to ions which strike the particle by virtue of their random heat motion. Previous analyses have assumed that the only field existing was that due to the charge on the particles, thus neglecting any external field which would be present in a corona discharge. Two effects of an external field which influence charging are critically examined: 1. Ions or electrons which travel in the direction of the external field toward the particle will acquire energy in addition to their heat energy. Their increase in energy will allow these elementary particles to overcome more readily the potential field of the charged particle. This effect is shown to be small for molecular ions at atmospheric pressure except for only the very smallest particles but, for charging by free electrons, which have a much longer mean free path, the effect could cause a large increase in charge. 2. The presence of an external field will alter the density distribution of ions around the charged particle since this field opposes the particle field on one side of the particle and aids it on the other side. The general equation for the ion density distribution was not able to be solved. However, approximate finite difference methods using computing machinery show that this effect is important and may almost completely account for the higher charge found with molecular ion charging.

THE CHARGING of particles by ions in a corona discharge has received erratic attention since 1923.[1–4] This work was reviewed and amplified by White[5] in 1951. Recent measurements[6,7] of the charge gained by fine particles in the presence of the strong electric fields which normally exist in a corona discharge have shown that the charge is appreciably higher than that given by the previous theoretical treatments. The discrepancy may be seen in Fig. 1 where the experimental results of Penney and Lynch are compared with the results of the theory given by Arendt and Kallmann, and by White.

The charging of fine particles is attributed almost completely to ions which strike the particle due to their random motion rather those which strike the particle due to their regular motion along the electric field lines. White has differentiated between these two charging mechanisms by calling the first "ion diffusion" or "random-motion" charging, and the latter "field" or "ion-bombardment" charging. An inspection of the previous theoretical treatments of diffusion charging[3,5] shows that the only field assumed existing in the neighborhood of the particle is that due to the charge on the particle. In normal usage, the field in the charging chamber of an electrostatic precipitator is comparable to the field at the surface of the particle due to the charge on the particle.

The purpose of this work is to determine the manner and magnitude of the effect of this external electric field on the random-motion particle-charging process. It will be shown that this field will have two main effects: 1. the effect on the energy of ions traveling toward the particle, and 2. the effect on the density distribution of ions in the neighborhood of the particle.

This paper, after presenting the general theory, will deal primarily with the results and not the details of the various methods used. This is because the methods are somewhat involved and a reasonable explanation would result in undue length. The reader who is interested in a more complete explanation of any part is referred to Murphy's thesis.[8] A brief listing of assumptions made throughout is given in the Appendix and also discussed more completely in reference 8.

General Approach to Determination of Particle Charge

The manner of treating the flow to the particle will follow lines which are closely

---

Paper 59-102, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted June 23, 1958; made available for printing December 17, 1958.

Arthur T. Murphy is with the University of Wichita, Wichita, Kans., and Felix T. Adler and Gaylord W. Penney are with the Carnegie Institute of Technology, Pittsburgh Pa.

This investigation was supported in part by an Aerotec-Thermix Fellowship and in part by the National Science Foundation. The work was done in part by Arthur T. Murphy in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Carnegie Institute of Technology. The authors wish to acknowledge the Mellon National Bank and Westinghouse Electric Corporation for their generous assistance in extending the use of their computing facilities.

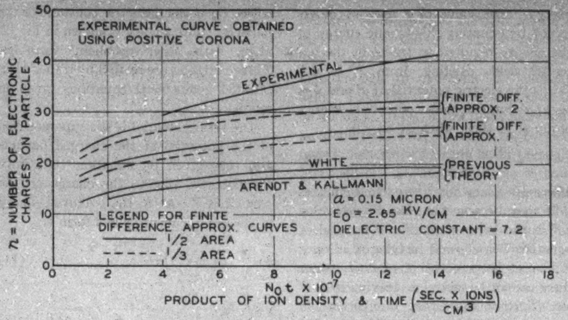


Fig. 1. Comparison of experimentally determined particle charge with previous theories and results of approximate ion density investigation

related to the treatment of diffusion in the kinetic theory.[9,10] A particle, introduced into a region of unipolar ions, of arbitrary density distribution with arbitrary electric fields present, will acquire charge due to ions which strike the particle. Once the particle is charged to some degree, the repelling action of this charge will cause a redistribution of ions in the immediate vicinity of the particle. The new density distribution of ions is then a function of the field of the particle superimposed on the previously existing fields.

An ion in space, heading for the particle, experiences a repelling force from the like-charged sphere. If the ion has a high enough velocity, it may still reach the particle over this repelling force. The number of ions which may do so is dependent on the number of ions which are at any point headed in the correct direction, with sufficient velocity, and which manage to avoid a collision with a molecule on their itinerary. All points in space must be considered, as well as all possible regions on the particle which ions may strike. Obviously, the further an ion must travel to reach the sphere, the less likely it is to avoid hitting a molecule. The counterbalancing of these effects is considered in the following mathematical treatment.

To determine the number of ions arriving at the entire particle surface per unit time, the rate of arrival of ions on an infinitesimal area dS in a general position on the sphere will first be considered (see Fig. 2), then an integration over the particle surface will give the desired result. The rate at which ions arrive at dS will be found by considering the number of ions which reach dS from an infinitesimal volume dV at a general point in space P, after having had a collision with a molecule at P. Then, an integration over all space that can "see" the area dS will yield the total rate of flow to dS. The expression "the volume dV at the point P" will be shortened to "the point P" in the following discussion.

Care must be taken that ions which may reach the particle are counted only once. This is guaranteed by the collision-type accounting. The only ions which are in dV and headed toward dS which are counted in the calculation will be those which have just had a collision in dV. Others which may be passing through P heading towards dS will be counted when P is further out (on the radius vector r) at the point where they had their last collision.

On the basis of the assumptions, the number of ions which have a collision at P per unit time with a speed between c and $c+dc$ after collision is

$$N\frac{c}{\lambda}\, f(c)dc\,dV \qquad (1)$$

where $f(c)$ is the distribution of speeds without regard to direction, i.e., $f(c)dc$ is the probability that an ion will have a speed between c and $c+dc$ after collision; N is the number of ions per unit volume at P. Of this number, the fraction $dS \cos\theta_r/4\pi r^2$ is heading in solid angle which dS makes with the point P if it is assumed that the scattering after collision is isotropic. Therefore, the number is

$$N\frac{c}{\lambda} f(c)dc\,\frac{dS\cos\theta_r}{4\pi r^2}\,dV \qquad (2)$$

ions per unit time with speeds between c and $c+dc$ leave P, after a collision, headed for dS.

Now, consider that only a fraction of those that leave P for dS will actually reach the area because: 1. there is a possibility that they may have a collision with a molecule before traveling the the distance r (and therefore should not be counted here, even if they eventually reach dS, since only those ions are being counted which reach dS after having had their last collision at P), and 2. the speed c may not be great enough to overcome the potential field due to the charge already on the particle.

Accordingly the result in expression 2 must be multiplied by the probability of not hitting a molecule in the distance r, i.e., $e^{-r/\lambda}$. Also, to count only those ions with velocity in the range such that they may reach the particle, it is necessary to integrate over all speeds above a certain minimum c_m, which is the speed an ion would need at P to just reach dS with zero velocity in the presence of the existing potential fields.

Hence, we now have the rate of arrival of ions at dS from P and if 100% attachment is assumed, this is also equal to the rate of increase of ions on the particle which have traveled from P to dS.

$$\left.\frac{dn}{dt}\right|_{P-dS} = N\,dV\,\frac{dS\cos\theta_r}{4\pi r^2}\,e^{-r/\lambda}\int_{c_m}^{\infty}\frac{c}{\lambda}\,f(c)dc \qquad (3)$$

An integration over all points that can be "seen" by dS gives with
$dV = r^2 \sin\theta_r d\theta_r d\phi_r dr$

r goes from 0 to ∞
θ_r goes from 0 to $\pi/2$
ϕ_r goes from 0 to 2π

$$\left.\frac{dn}{dt}\right|_{dS} = \frac{dS}{4\pi}\int_0^{\infty} dr \int_0^{\pi/2} d\theta_r \sin\theta_r \times$$
$$\int_0^{2\pi} d\phi_r\, N\cos\theta_r\, e^{-r/\lambda}\int_{c_m}^{\infty} c/\lambda\, f(c)dc \qquad (4)$$

where c_m is a function of position.

The form of equation 4 is such that it applies to any area dS on the sphere if the

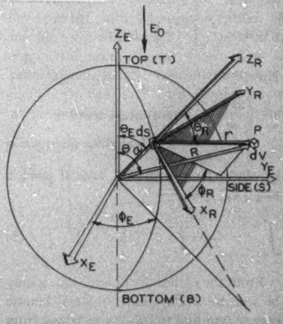


Fig. 2. Co-ordinate system for ion at point P in volume dV heading for area dS on sphere

quantities N and c_m are expressed in general form. The determination of the forms of the velocity (or energy) integral and N for the special case of a uniform external field will now be considered. The total charging rate may then be obtained by allowing dS to take on all values over the particle.

Energy Distribution of Ions

In the nonequilibrium situation of the ions in a force field, it is known that the velocity distribution function will deviate from that for an equilibrium situation i.e., the distribution function will not be exactly Maxwellian.[11] No general velocity distribution function for ions in a field exists in the literature. However, the distribution of velocities desired in equation 4 is the distribution of velocities just after collision. Upon collision, the ions, which have had a drift velocity in the electric field before collision, transfer some of their momentum and directed energy to the molecules. Then, in their next free path, they reattain the necessary drift velocity. The drift velocity immediately after collision will not be zero, but will be much less than the average drift velocity which ions have from one collision to the next. Consequently, assume that immediately after collision, the ions have completely "forgotten" that they were in an electric field, and the distribution of speeds may be taken as Maxwellian. Hence[12]

$$cf(c)dc=4\pi\left(\frac{m}{2\pi kT}\right)^{3/2}c^3\exp(-mc^2/2kT)dc \tag{5}$$

Defining

$$\xi=1/2\,\frac{mc^2}{kT} \tag{6}$$

yields

$$cf(c)dc=\frac{2}{\sqrt{\pi}}\left(\frac{2kT}{m}\right)^{1/2}\xi e^{-\xi}d\xi$$

Now,[13]

$$\frac{2}{\sqrt{\pi}}\left(\frac{2kT}{m}\right)^{1/2}=\bar{v}=\text{average speed of Maxwellian distribution}$$

so that $cf(c)dc=\bar{v}\xi e^{-\xi}d\xi$ and if $g=1/2\,mc_m^2/kT$, then

$$\int_{c_m}^{\infty}cf(c)dc=\bar{v}\int_{g}^{\infty}\xi e^{-\xi}d\xi \tag{7}$$

From the earlier definition of c_m, g may be seen to be the (dimensionless) kinetic energy required for an ion to travel from a point in space to the point on the particle at dS, just reaching dS with zero kinetic energy. The value of g may be determined by equating this kinetic energy to the energy required to travel in the potential fields existing, from P to dS. If ϕ is the electric potential at P and $\psi=e\phi/kT=$ dimensionless potential energy of ion at P, then, when $\psi_0-\psi\geq 0$

$$g=\psi_0-\psi \tag{8}$$

where $\psi_0=\psi$ at dS on the particle surface.

It may be seen that the lowest value which the lower limit of the integral of equation 7 is allowed to take on is zero; it cannot be a negative quantity since there cannot be more ions leaving from P for dS than the entire number which have had a collision at P and are headed for dS. Thus if the quantity $(\psi_0-\psi)$ should be negative, then g is replaced with zero in equation 7. The physical meaning of a negative g is evident from a consideration of the way in which the problem was developed. If an ion is moving against the field (assuming positive ions) it must have an initial velocity greater than zero to make any progress in the direction in which it started. If an ion first travels with the field and later against it, the energy which it gained in the first part of its trip will enable it to travel further through the repelling field. Or, stated in a different way, it may travel the same distance starting with a lower initial velocity. An ion which starts for the particle with zero velocity, and which travels a sufficient distance in the direction of the field, will have a residual velocity when it strikes the particle. Therefore, all ions headed towards dS which come from those regions in space where g is negative have sufficient velocity to reach the particle (although they may not actually reach the particle due to a collision). Thus

$$\bar{v}\int_{g}^{\infty}\xi e^{-\xi}d\xi=\begin{cases}\bar{v}(1+g)e^{-g} & \text{if } \psi_0-\psi\geq 0\\ \bar{v} & \text{if } \psi_0-\psi<0\end{cases} \tag{9}$$

The factor $(1+g)e^{-g}$ is thus seen to be the fraction of ions which have had a collision at P and have sufficient velocity to reach the particle. Fig. 3 shows a plot of this function versus g.

Minimum Ion Energy Under Influence of Uniform External Field

The problem of determining the potential in the neighborhood of a dielectric sphere in a uniform electric field is straight forward and may be found in any advanced text.[14] For the region outside the sphere the potential is

$$V=E_0\cos\theta\left[R-\frac{a^3}{R^2}\,\frac{(D_p-1)}{(D_p+2)}\right] \tag{10}$$

where D_p is the dielectric constant of the particle. The angle θ is shown in Fig. 2, along with other pertinent variables. These variables are related by rather complicated geometrical equations given in reference 8.

The total potential is the superposition of the potential of the charged particle and that given by equation 10. Since the potential due to the charge on the particle is $V=ne/R$, the potential energy of the ion in dimensionless form is

$$\psi=\frac{\gamma}{x}+w\left(x-\frac{\alpha}{x^2}\right)\cos\theta \tag{11}$$

$$x=\frac{R}{a},\quad w=\frac{eE_0a}{kT},\quad \alpha=\frac{D_p-1}{D_p+2},\quad \gamma=\frac{ne^2}{akT}$$

$$\psi_0=\gamma+w(1-\alpha)\cos\theta_E$$

The function $\psi_0-\psi$ is complicated and hence difficult to visualize, since it depends both on the position of the ion considered and the position of the element dS on the particle. Therefore it shall be examined here only for a few simple cases, namely the cases relative to the top T (i.e., considering dS to be situated on the top, where $\theta_E=0$); the bottom B ($\theta_E=\pi$); and the side S ($\theta_E=\pi/2$). For further analysis see reference 8. The set-up of the equations so far has implicitly assumed that the ions are positively charged, thus experiencing forces in the direction of the electric fields. The case of negative ions may be easily handled by changing the sign of e, the electronic charge. It is seen that the effect of doing so is to leave γ unchanged but to change w to $-w$. Thus, statements concerning θ_E will be reversed for negative ions, $\theta_E=0$ becoming $\theta_E=\pi$, and vice versa.

POTENTIAL ENERGY RELATIVE TO THE TOP

Consideration of the potential energy function relative to the top T shows that, at distances sufficiently far from the particle, an ion starting from rest gains energy from the external field to overcome the repelling field of the particle, and

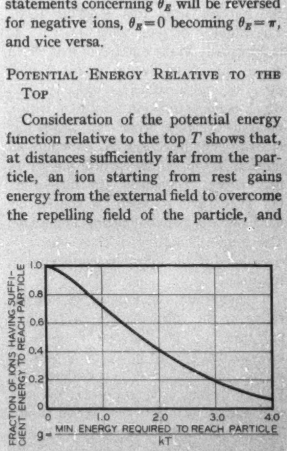


Fig. 3. Fraction of ions having sufficient energy to reach particle versus minimim ion energy required to reach particle

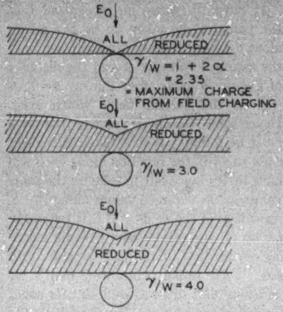


Fig. 4. Boundary of points for which all ions have sufficient energy to reach top (T) of particle. For various charge-to-field ratios

therefore $\psi_0-\psi<0$. This situation may be visualized by considering a surface representing the points from which an ion with zero speed at the boundary of the surface would reach dS with just zero speed. Ions which leave with zero velocity from points further from the particle than this surface will arrive at dS with a speed greater than zero.

The equation of this surface may be obtained by setting $g_{top}=0$, and the resulting relation defines the surface for any ϕ_r. The boundary can be drawn as shown in Fig. 4 for some typical values of γ/w with $\alpha=0.675$. For dioctylpthalate smoke whose dielectric constant is 7.2, $\alpha=0.675$. This smoke was used by the recent experimenters; see references 6 and 7. Analytical expressions and curves of the boundary of the sketches of Fig. 4 are presented in reference 8.

Those ions which leave from the region close to the particle marked "reduced" must have a velocity above some minimum in order to have a chance to reach dS; consequently only a fraction of the velocity distribution function is being counted, and this region will therefore be called the "reduced" region. The region on the other side of the boundary further from the particle is marked "all" since the entire distribution of velocities has a chance to reach dS if they leave from this region. The word "chance" is used here since there is still a probability that a collision with a molecule may occur between P and dS.

For 35 charges on an 0.15-micron radius particle ($\gamma=13.3$) and an external field of 5 kv/cm (per centimeter) ($w=2.85$), it is found that the boundary of the "all" region for $\theta_r=0$ is about 2.8 radii from the surface ($x=3.8, \theta_r=0$). This is the closest point at which all the ions have sufficient velocity to reach the particle. At 2.8 radii from the surface with zero external field, only 0.06% of the ions have speeds high enough to overcome the particle field. Thus the ratio of these is seen to be about 1,700 i.e., 1,700 times more ions have sufficient energy to reach the particle from a distance of 2.8 radii in the presence of a field of 5 kv/cm than do so without this field. Now, if the mean free path were long enough so that an appreciable number of these might survive on their trip to the particle, this effect would cause a large increase in the charging rate. For molecular ions at atmospheric pressure, the mean free path is such that only about 10^{-13} of the ions which are capable of reaching the particle, from 2.8 radii away, will actually arrive. Hence it may be seen that the increased energy effect will be small for molecular ions at atmospheric pressure.

A higher charge was found by Penney and Lynch for charging by means of negative corona at high wire potentials. They proposed that this could be due to the presence of some free electrons as well as negative ions at voltages near breakdown. It is instructive to look at the energy effect of the field on electrons. For electrons whose mean free path is much longer than ions, about 36% of the electrons can travel from 2.8 radii away without a collision, compared to 10^{-13} of the ions. Thus it is seen that the presence of the external field can produce an extremely significant increase in the charging rate from the ion energy effect when mean free paths of the order of the particle radius or larger are encountered. However, for usual molecular ions at atmospheric pressure, one must look for the major field effect to come from increased ion density close to the particle for all but the very smallest particles.

POTENTIAL ENERGY RELATIVE TO THE BOTTOM

The quantity g is always positive and it increases monotonically with distance from the particle. There are no boundaries, such as in Fig. 4, for the bottom B of the particle; g is always greater when an external field is present than it would be with no external field; therefore there will be fewer ions capable of reaching the particle from the bottom when the field is present.

POTENTIAL ENERGY RELATIVE TO THE SIDE

Relative to the side S, it may be determined that for $\pi/2<\phi_r<3\pi/2$, g may become negative while for $3\pi/2<\phi_r<\pi/2$, g is always positive. The region $\pi/2<\phi_r<3\pi/2$ is above the equator of the sphere, while the other region is below. This will then result in a boundary which can be pictured as in Fig. 5 for the plane $\phi_r=0, \pi$. For this case the effect of the field is to increase charging from the planes $\pi/2<\phi_r<3\pi/2$ and decrease it for $3\pi/2<\phi_r<\pi/2$.

SUMMARY OF EFFECT OF EXTERNAL FIELD ON ION ENERGY

1. Near the top of the sphere, ions heading toward the particle travel in the direction of external field, thereby gaining energy, and increasing the number of ions which have sufficient energy to reach the particle.
2. Near the bottom of the sphere, ions heading toward the particle travel against the field, thereby losing energy, and decreasing the number of ions which have sufficient energy to reach the particle.
3. Around the equator of the sphere, ions heading toward the particle from above travel with the main field, while ions from below travel against the main field. Thus it seems that the net effect on the side should be little different from the case of zero external field, which is reasonable since the total field is purely radial and equal to the particle field for $\theta_r=0$.

Effect of Electric Field on Ion Density

The ion density N in equation 4 must be known as a function of position in order that the integration be carried out. The particular form of the density distribution is determined by the particular fields which exist. However, a general equation for the ion density may be written for arbitrary fields.

The controlling partial differential equation for the ion density in a force field may be derived by considering the total net ion current and the conservation of matter. On the one hand, there is a simple diffusion of ions whose flow is proportional to the concentration gradient. On the other hand, there is a flow at every point proportional to, and in the

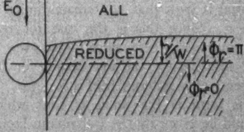


Fig. 5. Boundary of points for which all ions have sufficient energy to reach side (S) of particle

direction of, the electric field. When a field exists which attempts to disturb the density from a uniform distribution (such as the action of a charge on the particle because of the repelling action of this charge on the like-charged ions surrounding the particle), the diffusive flow comes into action in an attempt to return the distribution to uniform. In the steady state (if it is ever reached) there is a counterbalancing of these two effects which must meet whatever net flow conditions are prescribed by any external fields which may be present. This derivation is given in reference 8 along with the required boundary conditions. This equation is of the form given by Jost[15] as the equation of diffusion under the influence of external forces. To solve this exact equation with the specified boundary conditions is a formidable task. Approximations are also discussed in reference 8, where it is shown that it is reasonable to assume a steady state in time when solving for the ion density distribution and also to assume that there is zero flow of ions to the particle for purposes of determining the density. With these assumptions the diffusion equation for determining ion density is

$$\nabla^2 N + \nabla\psi \cdot \nabla N = 0 \tag{12}$$

$N \rightarrow N_0$ = ion density at infinite distance from particle as $x \rightarrow \infty$ for all θ

$$\text{grad}_x\, N + N\, \text{grad}_x\, \psi = \frac{\partial N}{\partial x} + \frac{\partial \psi}{\partial x}\, N = 0 \text{ at } x = 1$$

$$\nabla^2 N = \frac{1}{x}\frac{\partial^2}{\partial x^2}(xN) + \frac{1}{x^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial N}{\partial \theta}\right) \tag{13}$$

$$\nabla_x N = \frac{\partial N}{\partial x};\ \nabla_\theta N = \frac{1}{x}\frac{\partial N}{\partial \theta}$$

where

$$x = R/a;\quad \psi = \gamma/x + w(x - \alpha/x^2)\cos\theta$$

It may be seen that a solution of equation 12, which satisfies the boundary condition at the particle surface, is exponential $(-\psi)$. (This is true regardless of the form of ψ.) However, the boundary condition at infinity will be satisfied by this function only if the potential is such that $|\psi|$ remains finite as $x \rightarrow \infty$. Since this is the case for zero external field where, since $w = 0$, $\psi = \gamma/x$, the solution for the ion density in the absence of an external field is $N = N_0 e^{-\gamma/x}$. This form is the same as that which would be given by the Boltzmann factor from the kinetic theory of reference 16, and is the ion density used by White (only to evaluate the density at the particle surface, however). This result is to be expected since the steady state in the absence of an external field with no flow to the particle is an equilibrium situation and meets the conditions of the Boltzmann derivation.

Attempts at a mathematical solution of this problem for the potential given in equation 13 were unsuccessful. In the light of the availability of an IBM *650* (International Business Machines Corporation) computer and a small amount of time on an IBM *704*, a solution of equation 12 based on finite difference approaches was believed feasible. Many different attacks along the finite difference line were attempted but a complete solution was never attained. It was found that an extremely large number of points (resulting in a large number of simultaneous linear equations) would be required to obtain a solution of modest accuracy. These mathematical and finite difference approaches are described in detail in reference 8.

Due to the complexity of this problem it is necessary to build up evidence about the nature of the solution and the magnitude of the effect of the external field on the ion density, and therefore on the charging rate and final charge. The type of effect may be seen by considering that the charge on the particle attempts to cause a current flow outward from the particle while the external field attempts to cause a current flow toward the particle on the top, and away from the particle on the bottom. Hence, the external field should cause an increase in ion density at the top of the particle and a decrease at the bottom. An estimate of the magnitude of the effect will be made by means of finite difference approximations along $\theta = 0$, where the particle and external fields oppose each other; the major increase in density occurs there. The details of the mathematical basis for these approximations are given in reference 8; a brief description of the basic ideas of the method is given here.

Finite difference equations written for the points on the ray $\theta = 0$ are relations between the functional values of ion density along this ray and the functional values along the next closest ray (which will be designated ray k) in the direction of increasing θ. If values are known for the points along ray k, the form of the equation is such that a direct solution (not one by iteration) can be made for the unknowns along $\theta = 0$. Various values were assumed for the points on ray k with the intention of determining how much of an increase the external field would cause in the ion density near the surface of the particle at $\theta = 0$.

It must of course be recognized that the values to be taken on by the points along $\theta = 0$ are directly dependent on both the assumed values at k and the distance of the ray k away from the ray $\theta = 0$, as well as the various fields which exist. Hence some reasonable value for the ray k must be found. If it is taken extremely close to $\theta = 0$, the function along $\theta = 0$ will have little choice but to take on the same values as the assumed values along ray k, regardless of the fields. If it is taken too far away, the results may become meaningless because of the large angular-grid size. Fifty radial points were used after verifying that this was sufficient for small trunation error by comparing with results for 100 points. By means of a lengthy

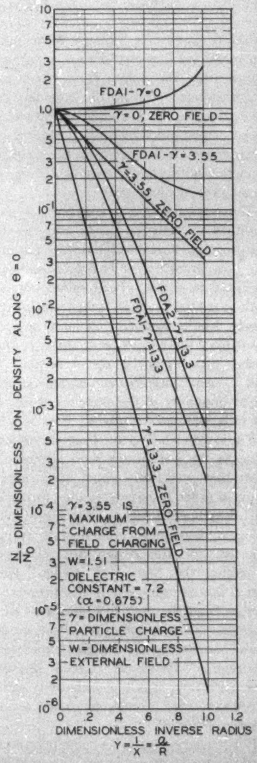


Fig. 6. Results of finite difference approximations for various particle charges. Typica case for $\gamma = 13.3$, $w = 1.51$: 35 charges on 0.15-micron radius particle in a field of 2.65 kv/cm

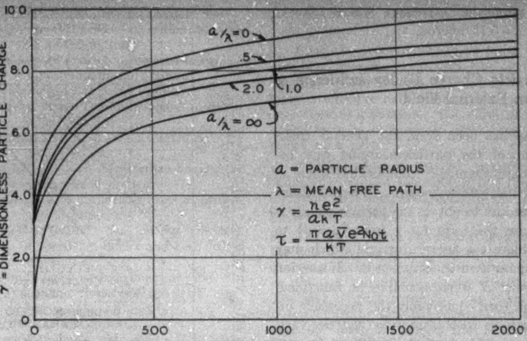


Fig. 7 Particle charge in absence of an external field for various ratios of particle radius to mean free path

series of trials, explained in reference 8, $\theta=\theta_k=25$ degrees was selected as a conservative value. If a θ greater than 25 degrees were chosen, the results for ion density would be higher.

Two solutions for two sets of values assumed along ray k are proposed. These will be denoted as FDA (finite difference approximations) 1 and 2. They are presented in an attempt to give limits for the density. The first, FDA 1, is found by assuming the values along ray k to be that which would exist at all angles in the absence of an external field, i.e., the values given by the Boltzmann factor $N_0e^{-\gamma/x}$. This value is known to be too low, and these results are proposed as a conservative estimate (conservative because $\theta_k=25$ degrees and not larger was used, and because $N_k=N_0e^{-\gamma/x}$ is an obviously lower value than actual) for a lower limit on the ion density along $\theta=0$ for an external field of the magnitude used.

FDA 2 is obtained as a conservative upper limit by allowing the values along ray k to be given by $N_k=N_0e^{-Ky}$, where K was varied from γ to zero (this is seen to be of the form of the Boltzmann factor but it really has no physical significance here, except for the case $K=\gamma$). As K is reduced the values along ray k become larger and, as would be expected, the values along $\theta=0$ also increase. For very small K, the values along $\theta=0$ become less than those along ray k, which is a physical impossibility. Hence the values found along $\theta=0$, when the surface value at $\theta=0$ is equal to the surface value at $\theta=\theta_k$, where taken as an upper limit and are referred to as FDA 2.

Fig. 6 shows the results of FDA 1 and 2 with a comparison with the zero field case for $\gamma=13.3$, $w=1.51$, $\alpha=0.675$. A typical case for these values is 35 charges on a 0.15-micron radius particle in an external field of 2.65 kv/cm. It may be seen that the factor by which the ion density at the particle surface at $\theta=0$ is increased due to the presence of the external field is between 130 and 440, as shown by the points for $y=1.0$ in Fig. 6.

It may be seen from the other curves of Fig. 6 that the increase over the Boltzmann factor caused by the field is a function of the charge on the particle as well as the external field. These curves are for the same external field but a charge on the particle equal to the saturation charge from field charging $\gamma=(1+2\alpha)w$ and a charge of zero determined by FDA 1. For the case of $\gamma=3.55$, $w=1.51$, we see that the field now causes an increase of only a factor of 5 over the Boltzmann factor compared to the lower limit of 130 for $\gamma=13.3$. For zero charge on the particle, the curve for $w=1.51$, $\gamma=0$, shows an increase over the Boltzmann factor (which is 1.0 everywhere for $\gamma=0$) as a factor of only 2.5. This shows that the factor of increase over the zero field value is greatly dependent on the particle charge as well as the external field present.

Lastly, one calculation was made at $\theta=180$ degrees to determine the order of magnitude by which the ion density is decreased by the field, which at $\theta=180$ degrees is in the same direction as the particle field. The results for FDA 1 (with the ray k now 25 degrees away from $\theta=180$ degrees) showed that the charging contribution from the lower part of the particle will be very small; a factor of decrease of 2.8 less than the zero field value being found. In addition, since the factor of decrease at $\theta=180$ degrees is far less than the factor of increase at $\theta=0$, the higher density demonstrated to be present at $\theta=0$ may exist over a considerable portion of the particle so that the average density may be considerably higher than just a simple average of the top and bottom densities would indicate. Since the approximations are based only on the behavior at $\theta=0$ and 180 degrees, one cannot more closely estimate this increase in average ion density.

SUMMARY OF EXTERNAL FIELD EFFECT ON ION DENSITY

From the previous discussion the following tentative conclusions may be drawn. The effect of the external field on the ion density is a very strong one. The density is increased over the zero field value where the external field opposes the particle field and decreased from the value where they are in the same direction. Results will be quoted for an external field of 2.65 kv/cm with 35 charges on a 0.15-micron particle of dielectric constant of 7.2, unless otherwise stated.

1. The lower limit of decrease at the surface at $\theta=180$ degrees is only about a factor of 2.8 less than the zero field value compared to a lower limit of increase at $\theta=0$ of 130. This indicates that the increase around $\theta=0$ may exist over an appreciable area.
2. The order of magnitude of increase at the surface at $\theta=0$ is conservatively from 130 to 440 times the zero field value.
3. The increase at $\theta=0$ varies with charge on the particle as well as with the external field. For an external field of 2.65 kv/cm the factor of increase over the zero field case at the surface given by the lower limit given is: 2.5 for an uncharged particle; 5 for 9 charges (which is the saturation charge from field charging); and 130 for 35 charges.

Particle Charge in Absence of an External Field

It is possible to perform a complete analytical solution for the zero field case with the use of equation 4 since N is given by the Boltzmann factor and g is sufficiently simplified that all integrations may be performed. The particle charge is found to depend on the mean free path (pressure effect) in a manner such as that shown in Fig. 7. The curve for zero mean free path ($a/\lambda=\infty$) is that given by White. This variation with pressure has not appeared in previous theories. The earlier discussion of the energy effect showed that the pressure variation will be larger when an external field is present although not paticularly significant for molecular ions at atmospheric pressure where $a/\lambda = 11$ for $a = 0.15$ micron.

Particle Charge Under Influence of an External Field

As has been seen, the general treatment of the particle-charging rate bebecomes extremely complicated when an external field is present. The necessary procedure to obtain the particle charge is to use the form for g from $(\psi_0 - \psi)$ to evaluate the energy integral which may then be used in equation 4. If the ion density N were available in functional form, then conceivably the necessary integrations with respect to the space variables could be performed with the form of the integrand changing whenever g went negative. Once this result is available, the total charging rate can be obtained by an integration over the surface of the particle

$$\frac{dn}{dt} = \int_S \frac{dn}{dt}\bigg|_{dS} = 2\pi a^2 \int_0^{\pi} d\theta_E \times \sin\theta_E \left\{ \frac{1}{dS} \frac{dn}{dt}\bigg|_{dS} \right\} \tag{14}$$

since $dS = a^2 \sin\theta_E d\theta_E d\phi_E$, where the integration with respect to ϕ_E gives 2π, since the integrand is independent of ϕ_E.

Certainly, because of its complexity, this procedure is to be avoided if possible. As will be remembered, the mean free path for ions is of the order of one tenth the particle radius, or less for particles of a 0.15-micron radius and larger. This leads to the conclusion that only those ions which have collisions very near the particle will be the ones that actually attach to the particle. That is, there is a little increase in the charging rate from the ion energy considerations, since ions come from very close to the particle, so they are able to acquire little increase in energy from the field. The important factor in determining the increased charging for molecular ions will be the ion density at the surface of the particle. (It is noted earlier that, for electrons, the energy effect is extremely important.) Hence it is proposed that the main effects for ions can be explained by calculations made on the basis of the limiting case of zero mean free path.

In the limit of zero mean free path, g becomes zero and the integrations may be performed giving

$$\frac{dn}{dt}\bigg|_{dS} = \frac{\bar{v} dS}{4} N_s \tag{15}$$

where N_S = ion density at the surface of the particle. Defining a dimensionless product of ion density and time as

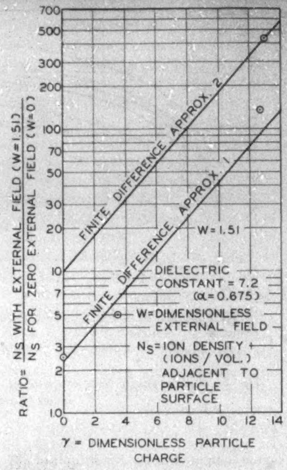


Fig. 8 Effect of external field on ion density at top of particle ($\theta = 0$) as function of particle charge. Typical case same as for Fig. 7

$$\tau = \frac{\pi a \bar{v} e^2 N_0 t}{kT}$$

gives the total charging rate as

$$\frac{d\gamma}{d\tau} = \frac{1}{2 N_0} \int_0^{\pi} d\theta_E \sin\theta_E N_s$$

This integral may be recognized as the average value of N_s over the surface of the particle, so

$$\frac{d\gamma}{d\tau} = \frac{N_{s(\text{avg})}}{N_0} \tag{16}$$

If it is assumed that about one half of the particle area has a surface ion density of the order of that at $\theta = 0$ and the other half has negligible charging, the average ion density over the particle surface is one half of that at $\theta = 0$. This factor of one half seems reasonable from earlier considerations, comparing the increase of density of $\theta = 0$ and the decrease at $\theta = 180$ degrees, which showed the lower limit of decrease at 180 degrees to be a factor of about 2.8, compared to a lower limit of increase at $\theta = 0$ of 130.

An approximate equation may be obtained from the results by a consideration of the factor of increase in ion density over the Boltzmann factor, as shown in Fig. 8, in which are plotted the factor of increase at $\theta = 0$ versus γ, the dimensionless charge on the particle for $w = 1.51$, and $\alpha = 0.675$ for the few tests which apply. It may be seen that the three points for FDA 1 tend to fall on a straight line on this semilogarithmic scale. The straight line is proposed as an estimate of the variation with charge of the factor of increase over the Boltzmann result for the particular values $w = 1.51$ and $\alpha = 0.675$. It may be seen from Fig. 1 that our prediction that the increase in ion density should be less for lower charges is in general agreement with the experimental results. For example at, say, 30 charges on the particle, the experimental rate is only a factor of 400 higher than the Boltzmann factor result, compared to the factor of 2,600 higher at 35 charges. Assuming the same sort of variation with γ for FDA 2 the upper curve of Fig. 8 may be obtained by drawing a line parallel to the lower line through the one data point at 440.

A straight line on the logarithmic plot is an exponential variation, of course, and the equations of the two lines are

$$\text{FDA } 2 = 10\, e^{+0.3\gamma}$$
$$\text{FDA } 1 = 2.2\, e^{+0.3\gamma}$$

and using the factor of one-half for the average ion density yields the following for the charging rate for $w = 1.51$, $\alpha = 0.675$:

$$\left(\frac{d\gamma}{d\tau}\right)_{\text{FDA2}} = 5e^{+0.3\gamma} e^{-\gamma} = 5e^{-0.7\gamma}$$

$$\left(\frac{d\gamma}{d\tau}\right)_{\text{FDA1}} = 1.1e^{+0.3\gamma} e^{-\gamma} = 1.1e^{-0.7\gamma} \tag{17}$$

A general equation of the form $d\gamma/d\tau = K_1 e^{-K_2\gamma}$ may be seen to integrate into

$$\gamma = \frac{1}{K_2} \ln(e^{K_2\gamma_0} + K_1 K_2 \tau) \tag{18}$$

where $\gamma = \gamma_0$ when $\tau = 0$.

The curves marked as FDA 1 and 2 in Fig. 1 were plotted from equation 18 with the use of K_1 and K_2 from equation 17. Also plotted in dashed lines are the respective FDA 1 and 2 obtained by assuming only one third of the particle area to have the increase at $\theta = 0$, and two thirds having negligible charging. It is seen that this assumption is not too critical since the 33% decrease in area from 1/2 to 1/3 caused only about a 4% decrease in charge at $N_0 t = 14 \times 10^{-7}$ and a 5% decrease at $N_0 t = 1 \times 10^{-7}$ for FDA 2.

Conclusions

It is seen that an increase in the ion density of the order of magnitude indicated here will explain the reason for the discrepancy between experiment and the previous theory for charging by positive corona at atmospheric pressure for particles in the size range around a 0.15-micron radius. This is particularly true when it is remembered that the increase may well be larger than is found from FDA 2 due to the conservative choice of $\theta_k = 25$ degrees. In addition, there will be a slight increase due to the finite but small mean free path which was called zero in approximation. It is believed that a more accurate determination of the ion density, perhaps by means of a complete finite difference solution, will bear out the general nature of the results stated here. The indications presented do show that a large effect exists due to the presence of the field; and this effect must be taken into account in some manner for a reasonably accurate determination of the particle charge.

In the case of charging by negative corona at higher wire potentials at atmospheric pressure, the results on the energy effect of the field show that if free electrons are present, even in an amount which is a small fraction of the number of negative ions, they may materially increase the particle charge. Little more than this may be said at present, since it is not known in what ratio free electrons would exist. Further work along this line would have to be done to indicate the magnitude of increase more accurately. The field effect on the density will be much less important for electron charging because of the longer free path which means that electrons come from distances at which the density is not very different from that at infinity.

The theory has the inherent capability of including variations with pressure by means of the variation of charge with mean free path of ions. The results of Fig. 7 for zero external field show this variation. This effect will be larger when the external field is considered and may be included if several numerical integrations are performed.

Nomenclature

Basic Quantities

a = particle radius, cm
n = number of electronic charges on particle
E_0 = uniform external field strength, statvolts/cm
e = electronic charge, 4.8×10^{-10}, statcoulombs
N_0 = ion density in absence of charged particles, ions/cc (cubic centimeters)
t = time measured from start of charging, sec (seconds)
D_P = dielectric constant of particle
T = absolute temperature, K (degrees Kelvin)
k = Boltzmann's constant = 1.381×10^{-16} erg/K
dV = infinitesimal volume at P, cm^3
R = distance from center of particle to P, cm
dS = infinitesimal area on particle, cm^2
x_r, y_r, z_r = Cartesian co-ordinates with dS as origin, cm
r, θ_r, ϕ_r = spherical co-ordinates for x_r, y_r, z_r system, cm, rad (radians)
N = ion density at point P, ions/cc
c = speed of ion (without regard to direction), cm/sec
λ = mean free path of ions colliding with molecules, cm
= 1.3×10^{-6} cm (indefinite) for ions at atmospheric pressure
= 4.1×10^{-5} cm for electrons at atmospheric pressure
$f(c)dc$ = probability that an ion will have a speed between c and $c+dc$ after collision
c_m = minimum speed which an ion can have at P and still reach dS, cm/sec
m = mass of ion, grams
$\bar{v}$ = average ion speed; 4.63×10^4 cm/sec for $T = 300$ K
x_E, y_E, z_E = Cartesian co-ordinates with center of sphere as origin, with E_0 parallel to z_E, cm
θ_E, ϕ_E = spherical co-ordinates for x_E, y_E, z_E system, rad
ϕ = electric potential, total, statvolts
V = electric potential, partial, statvolts
θ = angle between radius vector R and external field direction, rad
θ_k = value of θ for which the ion density was given assumed values, rad
N_s = ion density at the particle surface, ions/cc
$N_{s\,avg}$ = ion density averaged over the particle surface, ions/cc

Dimensionless Quantities

$\gamma = ne^2/akT$ = dimensionless (dml) charge on particle
γ_0 = initial dml charge on particle at time = zero
$w = eE_0a/kT$ = dml external field strength
$\alpha = (D_P - 1)/(D_P + 2)$
$\xi = mc^2/2kT$ = dml kinetic energy of ion with speed c
$g = mc_m{}^2/2kT$ = dml kinetic energy of ion with speed c_m
$\psi = e\phi/kT$ = dml electrical potential energy at P
$x = R/a$ = dml radius from center of sphere to P
$y = 1/x$ = dml inverse radius
N_k = assumed values of N along θ_k
$\tau = \pi a \bar{v} e^2 N_0 t/kT$ = dml time
K_1, K_2 = constants, see equation 18

Appendix. Assumptions

In all of the work here the following assumptions are made:

1. Particles are spherical.

2. Particle size is small compared to particle spacing.

3. The mean free path of ions is constant, independent of ionic velocity.

4. Ion-ion collisions are highly improbable due to the large difference in ionic and molecular densities (i.e., 10^9 ions/cc versus 10^{19} molecules/cc), therefore the mean free path of ions is independent of ionic density and depends only on molecular density, which is determined by pressure.

5. Charge is uniformly distributed on particles so that their fields are those of point charges.

6. Image forces (the attractive forces on individual ions due to the dielectric, or conducting, particle) are neglected. This force is effective only when an ion comes very close to a particle. The effect of the image force is to increase the charge attained by a particle.

7. All ions which strike the particle are assumed to remain attached to it. Image force considerations would indicate that this is a good assumption.

8. The space charge of ions traveling from the active to the passive electrode in the charging chamber causes the current to be limited and stabilizes the discharge. The field existing in the charging chamber is determined by this space charge, the wire potential and, of course, the geometry. In many practical cases the space charge is by far the predominant effect. Introduction of a particle (which soon becomes partially charged) will cause a local redistribution of the space charge.

Since the particle density is generally small compared to the ion density and the region of influence of a particle is limited to its immediate vicinity, the effect of this redistribution of the ion density on the field will be small. Consequently, it will be assumed that the external fields which a particle experiences are given by the value which exists in the absence of the particle.

9. As noted earlier, particle charging occurs by means of two mechanisms: field charging and random-motion charging. The field-charging results are given by White[6]; they show that for small particles and appreciable charging times that this charge is small compared to the final charge attained while a particle is in the charging chamber. Consequently, it is assumed that the particle acquires a charge equal to the maximum value from field charging very quickly and this value is used as the initial charge for calculations of charging by random motion. This maximum charge from field charging is given by

$$n = \frac{a^2E_0}{e}(1+2\alpha)$$

10. It will be assumed that ions travel in straight lines from collision to collision; i.e., the fields present cause negligible bending of the trajectories. This is partially justified by the fact that the majority of ions of interest travel nearly parallel to the field lines. In addition, any ions which do not hit the particle because of a curved path will have a mate, which, if it traveled in a straight line would miss the particle, but due to the curved path may well arrive at the particle. Thus there is a tendency for compensating effects due to curved paths. The inclusion of curved paths in the theory would result in extreme mathematical complication.

References

1. METHOD OF SIZE MEASUREMENT FOR SUSPENDED PARTICLES, H. Rohmann. *Zeitschrift fur Phvsik*, Berlin, Germany, vol. 17, 1923, pp. 253–65.

2. CHARGING OF SPHERICAL PARTICLES IN AN IONIZING FIELD, M. M. Pauthenier, M. Moreau-Hanot. *Journal de Phvsique et Le Radium*, Paris, France, series 7, vol. 3, 1932, pp. 590–613.

3. THE MECHANISM OF CHARGING OF MIST PARTICLES, **P. Arendt, H. Kallmann.** *Zeitschrift fur Physik*, vol. 35, 1925, pp. 421–41.

4. RESEARCH ON THE PHYSICAL BASIS OF ELECTRICAL GAS PURIFICATION, **R. Ladenburg.** *Annalen der Physik*, Leipsig, Germany, vol. 4, 1930, pp. 863–97.

5. PARTICLE CHARGING IN ELECTROSTATIC PRECIPITATION, **H. J. White.** *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1186–91.

6. MEASUREMENTS OF CHARGE IMPARTED TO FINE PARTICLES BY A CORONA DISCHARGE, **G. W. Penney, R. D. Lynch.** *Ibid.*, vol. 76, pt. I (*Communication and Electronics*), July 1957, pp. 294–99.

7. THE CHARGING OF SMALL PARTICLES FOR ELECTROSTATIC PRECIPITATION, **G. W. Hewitt.** *Ibid.*, pp. 300–06.

8. CHARGING OF FINE PARTICLES BY RANDOM MOTION OF IONS IN AN ELECTRIC FIELD, **A. T. Murphy.** *Ph.D. Thesis*, Carnegie Institute of Technology, Pittsburgh, Pa., Sept. 1956.

9. THE KINETIC THEORY OF GASES (book), **L. B. Loeb.** McGraw-Hill Book Company, Inc., New York, N. Y., 1934, pp. 252–77.

10. KINETIC THEORY OF GASES (book), **E. H. Kennard.** McGraw-Hill Book Company, Inc., 1938, pp. 184–205.

11. MOTION OF GASEOUS IONS IN STRONG ELECTRIC FIELDS, **G. H. Wannier.** *Bell System Technical Journal*, New York, N. Y., vol. 32, no. 1, Jan. 1946, pp. 170–254.

12. **E. H. Kennard,** *op. cit.*, p. 47.

13. **L. B. Loeb,** *op.cit.*, p. 105.

14. ELECTROMAGNETIC THEORY (book), **J. A. Stratton.** McGraw-Hill Book Company, Inc., 1941, p. 205.

15. DIFFUSION IN SOLIDS, LIQUIDS AND GASES (book), **W. Jost.** Academic Press, New York, N. Y., 1952, p. 46.

16. **E. H. Kennard,** *op. cit.*, p. 362.

Quality Assurance Program

R. A. HULNICK
NONMEMBER AIEE

H. G. HARDING
NONMEMBER AIEE

J. T. ROWINSKI
NONMEMBER AIEE

THE MANUFACTURE of components for a system as complex as the computer manufactured by IBM (International Business Machines Corporation) Kingston for use use in the SAGE (Semiautomatic Ground Environment) System requires strict adherence to specified requirements. To meet these requirements, analysis and control of vast numbers of fabricating operations are essential. These operations, although minor in themselves, can jeopardize the effective and specified function of the entire computer by their cumulative shortcomings.

Due to the size and complexity of the SAGE computer, which contains more than 7,371 pluggable units (each having approximately 450 solder connections), 50,000 vacuum tubes, 170,000 diodes, and 600,000 resistors, the problem of developing quality assurance procedures applicable to the manufacture of these components is critical.

IBM has conducted intensive studies to assure that the manufacture of components for complex systems such as SAGE matches specified requirements. Results have shown that the standards of workmanship in all areas are being raised to meet the required quality levels by using a quality assurance procedure which employs the following fundamental principles:

1. Selection of a numerical basis for quality rating.
2. Classification of defects according to importance in component function and area of incidence.
3. Determination of the percentage limits of acceptability.
4. Planned audit inspection.
5. Planned master control.

The quality assurance procedure presently used by the quality engineering department of IBM Kingston is outlined in this report. The procedure is based upon the use of four charts on which the information listed in Table I is compiled and displayed.

It is the intent of this report to present a program which will: 1. develop a system for examining and rating each electromechanical component-manufacturing area and each product of these areas prior to sending the component to stock, and 2. determine the inspection efficiency for each inspection.

The main topics which are discussed are master control of manufacturing areas and audit control of material in stock. Interrelation of these two (including the inspection efficiency nomograph) is described separately. Derivations of all formulas are given in the appendixes.

Master Control of Manufacturing Areas

THEORY

To establish the master control system, the first consideration was to obtain a common denominator for the evaluation of all manufacturing areas. Because some manufacturing areas produce more complex subassemblies than others, a denominator had to be determined which would be equitable and applicable to all areas. This denominator is a complexity factor based upon the total number of solder joints produced by a manufacturing area; i.e., the more complex a subassembly, the more solder joints that subassembly will have, and from this, the denominator is determined.

The second consideration, also fundamental to the master control system, was the determination of the per cent-defective (based upon the total number of solder joints) allowable for each classification of defects. The classification of defects and the per-cent-defectives allowable in the end product as shown in Table II are the result of a special study initiated by the IBM Kingston quality engineering department.

Major solder defects include:

1. No solder at the joint.

Table I. Quality Assurance Charts

Name of Chart	Purpose	Information Displayed
Per-cent-defective	Reveals trouble areas of each classification of defects. A chart for each manufacturing area makes production supervisor aware of conformance of area to limits	Trend lines show the per-cent-defective within each defect class with respect to each limit line, including number and description of the most prevalent defects
Master rating control	Examines and rates each manufacturing area according to actual quality achieved	The per-cent-defective and quality rating of types of defects for each manufacturing area as disclosed by standard inspection procedures
Audit rating control	Examines and rates the product as sent to stock	The per-cent-defective and quality rating of types of defects passed into stock as disclosed by audit inspection
Inspection efficiency nomograph	Determines the inspection efficiency of each inspection area	Inspection efficiencies with respect to master control rating for each inspection area

Vol.
Pt. 1
1959

Paper 59-144, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted March 18, 1958; made available for printing December 1, 1958.

R. A. HULNICK, H. G. HARDING, and J. T. ROWINSKI are with International Business Machines Corporation, Kingston, N. Y.

Table II. Classification of Defects and Allowable Per-Cent-Defective

Defect Classification	Allowable Per-Cent-Defective, %
Major solder	0.0075
Other major	0.01
Minor	0.03

2. Cold solder (crystallized and without feathering at the edges).
3. Loose solder (one or more wires unsecured and free to creep along the terminal).
4. Solder short (a tail or bridge connection between two conductors; not common).

Other major defects include damaged components, sheared wire strands, and components improperly mounted or improperly located.

Minor defects include wire dress, burrs, and unsatisfactory surface dress.

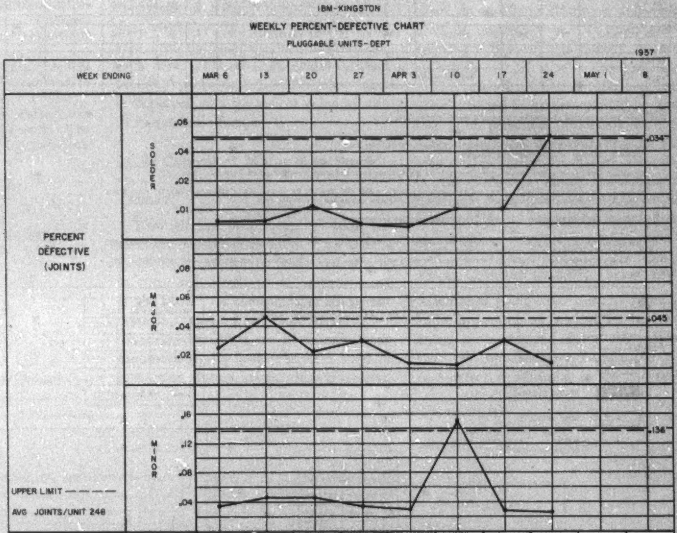
These defects will be referred to as solder, major, and minor, respectively, throughout the remainder of this report.

The maximum per-cent-defective allowable in a manufacturing area to assure an end product of acceptable quality can be calculated, given the basic per cent-defective allowable in the end product (Table II), and assuming a minimum regular inspection efficiency of 82%. See Table III for the maximum allowable per-cent-defective in a manufacturing area and Appendix I for the derivation of the formula for that per-cent-defective.

The final consideration in planning the master control system was the establishment of a method whereby each manufacturing area could be rated with respect to the quality requirements of the end product. Quality ratings are based on a

Table III. Maximum Allowable Per-Cent-Defective for Manufacturing Areas

Defect Classification	Maximum Allowable Per-Cent-Defective, %
Solder	0.034
Major	0.045
Minor	0.136



			MAR 6	13	20	27	APR 3	10	17	24	MAY 1	8
SOLDER		PERCENT DEFECTIVE	.006	.006	.012	.005	.004	.010	.010	.035		
MAJOR		PERCENT DEFECTIVE	.024	.046	.022	.030	.014	.013	.030	.014		
MINOR		PERCENT DEFECTIVE	.035	.047	.046	.035	.028	.151	.025	.023		
DESCRIPTION AND NO. OF DEFECTS	SOLDER	NO SOLDER	7	6	12	3	4	6	10	5		
		LOOSE SOLDER	2	1	2	1		1	3	53		
		OTHER SOLDER	2	2	3	4	2	6	6	1		
	MAJOR	BROKEN STRANDS	13	19	25	43	13	8	11	5		
		DAMAGED COMPONENTS	12	8	3	6	9	5	7	7		
		COMPONENT MOUNTING	4	29	2				1	6		
		IMPROPER COMPONENT	5	16				3		1		
		OTHER MAJOR	7		2				37	5		
	MINOR	INSULATION	15	21	22	19	19	10	9	7		
		INSUFFICIENT SOLDER		6	3	3	7	6	7	2		
		MECHANICAL WIRE CON'CNS	22	31	20	23	11	31	11	6		
		OTHER MINOR	22	16	21	12	7	143	19	24		
TOTAL NUMBER OF DEFECTS			111	155	115	114	72	219	121	122		
TOTAL UNITS INSPECTED			684	631	578	655	636	508	752	680		

Fig. 1. Weekly per-cent-defective chart distributed weekly to all managers involved with particular area. Top of chart illustrates per-cent-defective trend lines of solder, major, and minor characteristics in comparison to respective P_t limits. Center lists weekly per-cent-defective numerically. Bottom lists number of most numerous and persistent defects in each classification. As prevalent defects are reduced to insignificant levels, they are moved to miscellaneous group and other defects take their place

MASTER RATING WORK SHEET WEEK ENDING: APRIL 24, 1957

① AREA	② JOINTS PER UNIT	③ NO. OF UNITS INSP.	④ TOTAL JOINTS	⑤ NO OF DEFECTS				⑥ PERCENT DEFECTIVE (JOINTS)			⑦ 100-RATING FACTORS			⑧ MASTER RATING			
				SO	MA	MI	TOTAL	SO	MA	MI	530 SO	400 MA	132 MI	SO	MA	MI	TOTAL
PLUGGABLE UNITS	248	680	168,640	59	24	39	122	.035	.014	.023	19	6	3	81	94	97	91
BACK PANELS DEPARTMENT 215	1000	11	11,000	2	10	11	23	.018	.091	.100	10	36	13	90	64	87	80
BACK PANELS DEPARTMENT 216	1000	10	10,000	0	11	2	13	0	.110	.020	0	44	3	100	56	97	84
PLUGGABLE UNIT ASSEMBLY	☒	☒	189,640	61	45	52	158	.032	.024	.027	17	10	4	83	90	96	90
DEPARTMENT 210	256	72	18,432	7	8	48	63	.038	.043	.260	20	17	34	80	83	66	76
PRINTED CIRCUITS CARDS	20*	874	17,480	—	14	5	19	—	.080	.029	—	32	4	—	68	96	82
CARD ASSEMBLY DEPARTMENTS 205-6	26	5632	146,432	66	54	23	143	.045	.037	.016	24	15	2	76	85	98	86
DEPARTMENT 207	40	258	10,320	1	1	0	2	.010	.010	0	5	4	0	95	96	100	97
PLUGGABLE UNIT SUB-ASSEMBLY	☒	☒	156,752(SO) 174,232(MA-MI)	67	69	28	164	.043	.040	.016	23	16	2	77	84	98	86
TOTAL PLUGGABLE UNIT AREA	☒	☒	364,824(SO) 382,304(MA-MI)	135	122	128	385	.037	.032	.033	20	13	4	80	87	96	88
FRAMES DEPARTMENTS 221-2	**	8	9,280	1	4	15	20	.011	.043	.162	6	17	21	94	83	79	85
FRAMES DEPARTMENT 223	**	3	3,110	0	1	2	3	0	.032	.064	0	13	8	100	87	92	93
FINAL ASSEMBLY	☒	☒	12,390	1	5	17	23	.008	.040	.137	4	16	18	96	84	82	87
TOTAL MANUFACTURING AREA	☒	☒	377,214(SO) 394,694(MA-MI)	136	127	145	408	.036	.032	.037	19	13	5	81	87	95	88

Fig. 2. Master rating work sheet

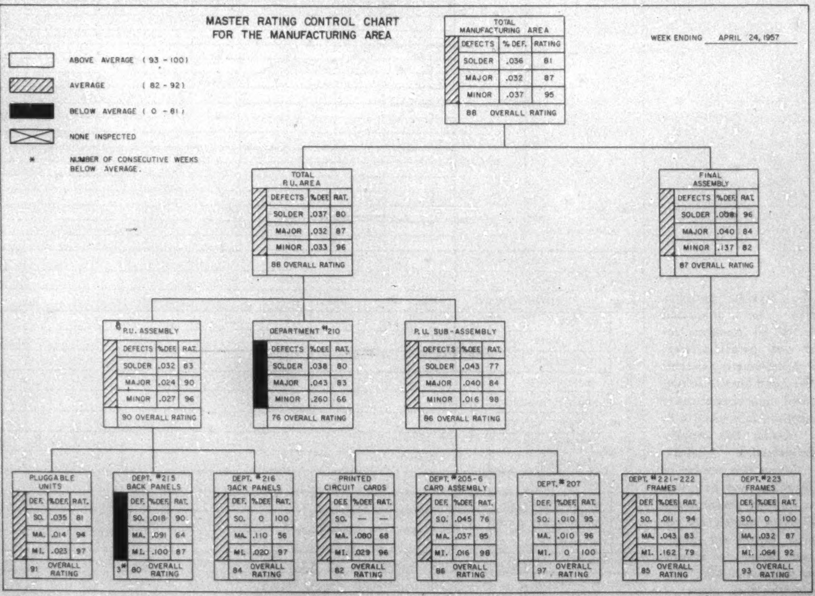
Procedure for calculating the master rating of a manufacturing area.

Column 1 - Area of manufacture.
Column 2 - Number of joints (complexity base) assigned to a particular unit.
Column 3 - Number of units inspected (weekly).
Column 4 - ② X ③
Column 5 - Number of solder, major, and minor defects per week as recorded from Weekly Percent-Defective Chart (FIG. 1).
Column 6 - ⑤ ÷ ④ X 100
Column 7 - Factor (constant) X ⑥. The factors (constants) used are 530, 400, and 132 for solder, major and minor defects respectively.
Column 8 - For solder, major, and minor ratings: 100-⑦. Negative values are interpreted as 0.

For overall rating: $100-\frac{\Sigma ⑦}{3}$ Negative values are interpreted as 0.

* Printed circuit cards are not soldered. A complexity number of 20 was assigned as a measure of unit complexity as related to those units whose complexity is based on the number of solder joints in the unit.
**Due to the great variations of complexity between the different types of frames, complexity figures were determined for each type of unit.

Fig. 3 (below). Master rating control chart for manufacturing area


MASTER RATING CONTROL CHART FOR THE MANUFACTURING AREA
WEEK ENDING APRIL 24, 1957
ABOVE AVERAGE (93 - 100)
AVERAGE (82 - 92)
BELOW AVERAGE (0 - 81)
NONE INSPECTED
* NUMBER OF CONSECUTIVE WEEKS BELOW AVERAGE.
TOTAL MANUFACTURING AREA
DEFECTS % DEF. RATING
SOLDER .036 81
MAJOR .032 87
MINOR .037 95
88 OVERALL RATING
TOTAL P.U. AREA
DEFECTS %DEF. RAT.
SOLDER .037 80
MAJOR .032 87
MINOR .033 96
88 OVERALL RATING
FINAL ASSEMBLY
DEFECTS %DEF. RAT.
SOLDER .008 96
MAJOR .040 84
MINOR .137 82
87 OVERALL RATING
P.U. ASSEMBLY
DEFECTS %DEF. RAT.
SOLDER .032 83
MAJOR .024 90
MINOR .027 96
90 OVERALL RATING
DEPARTMENT #210
DEFECTS %DEF. RAT.
SOLDER .038 80
MAJOR .043 83
MINOR .260 66
76 OVERALL RATING
P.U. SUB-ASSEMBLY
DEFECTS %DEF. RAT.
SOLDER .043 77
MAJOR .040 84
MINOR .016 98
86 OVERALL RATING
PLUGGABLE UNITS
DEF. %DEF. RAT.
SO. .035 81
MA. .014 94
MI. .023 97
91 OVERALL RATING
DEPT. #215 BACK PANELS
DEF. %DEF. RAT.
SO. .018 90
MA. .091 64
MI. .100 87
3* 80 OVERALL RATING
DEPT. #216 BACK PANELS
DEF. %DEF. RAT.
SO. 0 100
MA. .110 56
MI. .020 97
84 OVERALL RATING
PRINTED CIRCUIT CARDS
DEF. %DEF. RAT.
SO. — —
MA. .080 68
MI. .029 96
82 OVERALL RATING
DEPT. #205-6 CARD ASSEMBLY
DEF. %DEF. RAT.
SO. .045 76
MA. .037 85
MI. .016 98
86 OVERALL RATING
DEPT. #207
DEF. %DEF. RAT.
SO. .010 95
MA. .010 96
MI. 0 100
97 OVERALL RATING
DEPT. #221-222 FRAMES
DEF. %DEF. RAT.
SO. .011 94
MA. .043 83
MI. .162 79
85 OVERALL RATING
DEPT. #223 FRAMES
DEF. %DEF. RAT.
SO. 0 100
MA. .032 87
MI. .064 92
93 OVERALL RATING

scale of 0–100 and are divided into three general categories:

100–93 = area above average
92–82 = area average
81–0 = area below average

Operational Procedure

The operational procedure covering the master control of manufacturing areas comprises the following steps:

1. Inspection results are reported on a weekly basis to the quality engineering department. This weekly chart (Fig. 1) describes each manufacturing area according to the following considerations:
   (*a*). The number of units inspected.
   (*b*). The number and description of solder defects, major defects, and minor defects found in each manufacturing area.
   (*c*). Trend lines of solder per-cent-defectives, major defect per-cent-defectives, and minor defect per-cent-defectives in relation to their respective limit lines.
2. Information accumulated on the weekly quality control chart is then transferred to the master rating work sheet (Fig. 2). Using this work sheet, quality ratings are derived for each manufacturing area. The formulas (Appendix II) used for these calculations are as follows:

$$\text{Master rating (solder)} = 100 - \frac{53{,}000}{N}\, d_s$$

$$\text{Master rating (major)} = 100 - \frac{40{,}000}{N}\, d_{ma}$$

$$\text{Master rating (minor)} = 100 - \frac{13{,}200}{N}\, d_{mi}$$

$$\text{Master rating (over-all)} = 100 - \frac{4{,}400}{N} \times (4d_s + 3d_{ma} + d_{mi})$$

where

N = predetermined base value of the manufactured unit multiplied by the number of units inspected during one week
d_s = total number of solder defects found by regular inspection procedures (for one week)
d_{ma} = total number of major defects found by regular inspection procedures (for one week)
d_{mi} = total number of minor defects found by regular inspection procedures (for one week)

Chart Displays

Quality ratings calculated according to the operational procedure described are then posted on the master rating chart for the manufacturing area (Fig. 3). This chart shows the actual per-cent-defectives of solder, major, and minor characteristics for each manufacturing area and the quality ratings given for each area. The over-all quality rating of each particular manufacturing area is displayed as a numerical value as well as in symbolic form.

Audit Control of Material in Stock

Theory

Through the proper use of audit inspection, regular inspection of the manufacturing areas can be improved which, in turn, will result in an improvement of the quality of the manufactured components. Audit inspection teams cover all inspection areas on a periodic basis to ascertain the best type and level of inspection for a given manufacturing area. Particular attention is given to areas with below-average ratings. See Fig. 3 for a typical list of areas visited. The specific objectives of the audit inspection activity are as follows:

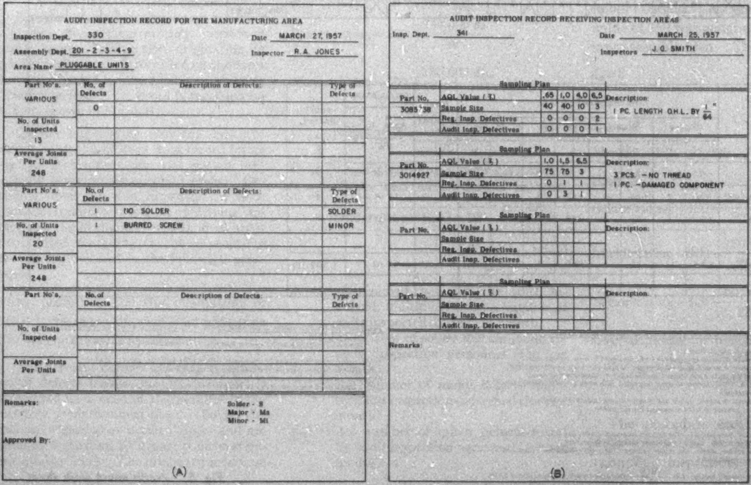
1. To determine if the proper tools and procedures are being used by regular inspection.
2. To recommend improved methods of inspection when current methods are found to be inadequate.
3. To help determine the proper frequency of inspection.
4. To make available a written description of the condition of the completed product of a given manufacturing area.
5. To aid in determining the product characteristics most in need of inspection.

The audit inspection activity is, in effect, an additional assurance that a high level of inspection will be maintained, assuring a flow of high-quality material into stock.

The type of inspection (100% or sampling) utilized by an area visited by audit inspection personnel provides a basis for determining the quality rating of that area. This arrangement is described in the following under two separate headings: areas covered by 100% inspection, and areas covered by sampling inspection.

Areas Covered by 100% Inspection

Material going into stock is rated by audit inspection as a weighted per-cent-defective subtracted from 100. When that per-cent-defective equals the per-cent-defective allowable in the end prod-



AUDIT INSPECTION RECORD FOR THE MANUFACTURING AREA

Inspection Dept. 330
Date MARCH 27, 1957
Assembly Dept. 201 - 2 - 3 - 4 - 9
Inspector R. A. JONES
Area Name PLUGGABLE UNITS

Part No's. VARIOUS	No. of Defects	Description of Defects:	Type of Defects
	0		
No. of Units Inspected 13			
Average Joints Per Units 248			
Part No's. VARIOUS	No. of Defects	Description of Defects:	Type of Defects
	1	NO SOLDER	SOLDER
No. of Units Inspected 20	1	BURRED SCREW	MINOR
Average Joints Per Units 248			
Part No's.	No. of Defects	Description of Defects:	Type of Defects
No. of Units Inspected			
Average Joints Per Units			

Remarks:
Solder - S
Major - Ma
Minor - Mi

Approved By:

(A)

AUDIT INSPECTION RECORD RECEIVING INSPECTION AREAS

Insp. Dept. 341
Date MARCH 25, 1957
Inspectors J. G. SMITH

	Sampling Plan					
Part No. 3085738	AQL Value (%)	.65	1.0	4.0	6.5	Description: 1 PC. LENGTH O.H.L. BY $\frac{1}{64}$"
	Sample Size	40	40	10	3	
	Reg. Insp. Defectives	0	0	0	2	
	Audit Insp. Defectives	0	0	0	1	

	Sampling Plan					
Part No. 3014927	AQL Value (%)	1.0	1.5	6.5		Description: 3 PCS. – NO THREAD 1 PC. – DAMAGED COMPONENT
	Sample Size	75	75	3		
	Reg. Insp. Defectives	0	1	1		
	Audit Insp. Defectives	0	3	1		

	Sampling Plan					
Part No.	AQL Value (%)					Description:
	Sample Size					
	Reg. Insp. Defectives					
	Audit Insp. Defectives					

	Sampling Plan					
Part No.	AQL Value (%)					Description:
	Sample Size					
	Reg. Insp. Defectives					
	Audit Insp. Defectives					

Remarks:

(B)

Fig. 4. Audit inspection record

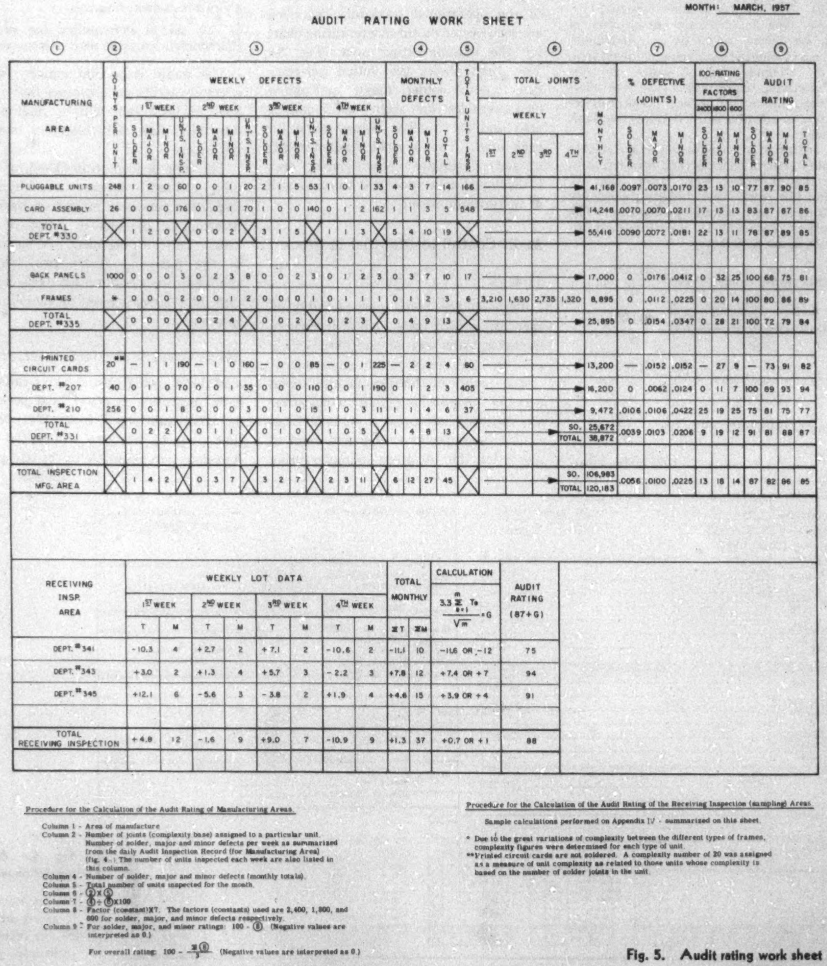
A—For manufacturing area
B—For receiving inspection areas

uct, the material is considered average and is given a numerical rating of 87, ±5. Since the range of this average rating has the same upper and lower limits as the master control chart (Fig. 3), it is possible to make comparisons between the master control and audit rating control charts in the 100% areas.

Areas Covered by Sampling Inspection

Material going into stock is rated as average when the per-cent-defective found by audit inspection personnel is the same as the per-cent-defective found by regular inspection personnel. Because of sampling risks, however, it cannot be assumed that regular inspection will always find a per-cent-defective that is equal to the per-cent-defective found by audit inspection. Accordingly, some deviation from the average value must be permitted. In this instance a ±1.5 sigma spread is allowed for sampling risks, incorporating a 0.134 probability that material which receives a rating outside the average range is, in fact, average in quality. Using the ±1.5 sigma spread, the low limit of the average range is 82, while the high limit is 92.

Ratings acquired through either method previously described are calculated on the proper work sheet and are then posted on the audit rating control chart for sampling areas. This chart shows the quality of material being passed into stock. Information appearing on the chart is displayed in two ways: 1. symbolically, as above average, average, or below average



MONTH: MARCH, 1957

AUDIT RATING WORK SHEET

①	②	③																④				⑤	⑥					⑦			⑧			⑨			
MANUFACTURING AREA	JOINTS PER UNIT	WEEKLY DEFECTS																MONTHLY DEFECTS				TOTAL UNITS INSP.	TOTAL JOINTS					% DEFECTIVE (JOINTS)			100-RATING FACTORS			AUDIT RATING			
		1ST WEEK				2ND WEEK				3RD WEEK				4TH WEEK									WEEKLY				MONTHLY				2400	1800	600				
		SOLDER	MAJOR	MINOR	UN'TS. INSP.	SOLDER	MAJOR	MINOR	UN'TS. INSP.	SOLDER	MAJOR	MINOR	UN'TS. INSP.	SOLDER	MAJOR	MINOR	UN'TS. INSP.	SOLDER	MAJOR	MINOR	TOTAL		1ST	2ND	3RD	4TH		SOLDER	MAJOR	MINOR	SOLDER	MAJOR	MINOR	SOLDER	MAJOR	MINOR	TOTAL
PLUGGABLE UNITS	248	1	2	0	60	0	0	1	20	2	1	5	53	1	0	1	33	4	3	7	14	166	—	—	—	→	41,168	.0097	.0073	.0170	23	13	10	77	87	90	85
CARD ASSEMBLY	26	0	0	0	176	0	0	1	70	1	0	0	140	0	1	2	162	1	1	3	5	548	—	—	—	→	14,248	.0070	.0070	.0211	17	13	13	83	87	87	86
TOTAL DEPT. #330	X	1	2	0	X	0	0	2	X	3	1	5	X	1	1	3	X	5	4	10	19	X	—	—	—	→	55,416	.0090	.0072	.0181	22	13	11	78	87	89	85
BACK PANELS	1000	0	0	0	3	0	2	3	8	0	0	2	3	0	1	2	3	0	3	7	10	17	—	—	—	→	17,000	0	.0176	.0412	0	32	25	100	68	75	81
FRAMES	*	0	0	0	2	0	0	1	2	0	0	0	1	0	1	1	1	0	1	2	3	6	3,210	1,630	2,735	1,320	8,895	0	.0112	.0225	0	20	14	100	80	86	89
TOTAL DEPT. #335	X	0	0	0	X	0	2	4	X	0	0	2	X	0	2	3	X	0	4	9	13	X	—	—	—	→	25,895	0	.0154	.0347	0	28	21	100	72	79	84
PRINTED CIRCUIT CARDS	20**	—	1	1	190	—	1	0	160	—	0	0	85	—	0	1	225	—	2	2	4	60	—	—	—	→	13,200	—	.0152	.0152	—	27	9	—	73	91	82
DEPT. #207	40	0	1	0	70	0	0	1	35	0	0	0	110	0	0	1	190	0	1	2	3	405	—	—	—	→	16,200	0	.0062	.0124	0	11	7	100	89	93	94
DEPT. #210	256	0	0	1	8	0	0	0	3	0	1	0	15	1	0	3	11	1	1	4	6	37	—	—	—	→	9,472	.0106	.0106	.0422	25	19	25	75	81	75	77
TOTAL DEPT. #331	X	0	2	2	X	0	1	1	X	0	1	0	X	1	0	5	X	1	4	8	13	X	—	—	→	SO. TOTAL	25,672 38,872	.0039	.0103	.0206	9	19	12	91	81	88	87
TOTAL INSPECTION MFG. AREA	X	1	4	2	X	0	3	7	X	3	2	7	X	2	3	11	X	6	12	27	45	X	—	—	→	SO. TOTAL	106,983 120,183	.0056	.0100	.0225	13	18	14	87	82	86	85

RECEIVING INSP. AREA	WEEKLY LOT DATA								TOTAL MONTHLY		CALCULATION	AUDIT RATING (87+G)
	1ST WEEK		2ND WEEK		3RD WEEK		4TH WEEK				$\frac{3.3 \sum_{\theta=1}^{m} T_\theta}{\sqrt{m}} = G$	
	T	M	T	M	T	M	T	M	ΣT	ΣM		
DEPT. #341	-10.3	4	+2.7	2	+7.1	2	-10.6	2	-11.1	10	-11.6 OR -12	75
DEPT. #343	+3.0	2	+1.3	4	+5.7	3	-2.2	3	+7.8	12	+7.4 OR +7	94
DEPT. #345	+12.1	6	-5.6	3	-3.8	2	+1.9	4	+4.6	15	+3.9 OR +4	91
TOTAL RECEIVING INSPECTION	+4.8	12	-1.6	9	+9.0	7	-10.9	9	+1.3	37	+0.7 OR +1	88

Procedure for the Calculation of the Audit Rating of Manufacturing Areas.

Column 1 - Area of manufacture
Column 2 - Number of joints (complexity base) assigned to a particular unit.
Number of solder, major and minor defects per week as summarized from the daily Audit Inspection Record (for Manufacturing Area) (fig. 4.) The number of units inspected each week are also listed in this column.
Column 4 - Number of solder, major and minor defects (monthly totals).
Column 5 - Total number of units inspected for the month.
Column 6 - ②X⑤
Column 7 - ④÷⑥X100
Column 8 - Factor (constant)X7. The factors (constants) used are 2,400, 1,800, and 600 for solder, major, and minor defects respectively.
Column 9 - For solder, major, and minor ratings: 100 - ⑧. (Negative values are interpreted as 0.)

For overall rating: $100 - \frac{\Sigma ⑧}{3}$ (Negative values are interpreted as 0.)

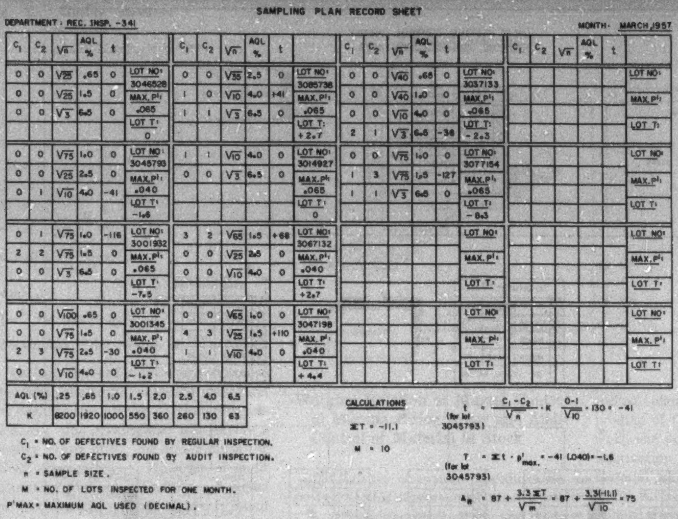
Procedure for the Calculation of the Audit Rating of the Receiving Inspection (sampling) Areas.

Sample calculations performed on Appendix IV - summarized on this sheet.

* Due to the great variations of complexity between the different types of frames, complexity figures were determined for each type of unit.

** Printed circuit cards are not soldered. A complexity number of 20 was assigned as a measure of unit complexity as related to those units whose complexity is based on the number of solder joints in the unit.

Fig. 5. Audit rating work sheet



SAMPLING PLAN RECORD SHEET

DEPARTMENT: REC. INSP. -341

MONTH: MARCH, 1957

C_1	C_2	$\sqrt{n}$	AQL %	t		C_1	C_2	$\sqrt{n}$	AQL %	t		C_1	C_2	$\sqrt{n}$	AQL %	t		C_1	C_2	$\sqrt{n}$	AQL %	t	
0	0	$\sqrt{25}$	.65	0	LOT NO: 3046528	0	0	$\sqrt{35}$	2.5	0	LOT NO: 3085738	0	0	$\sqrt{40}$	.65	0	LOT NO: 3037133						LOT NO:
0	0	$\sqrt{25}$	1.5	0	MAX. P': .065	1	0	$\sqrt{10}$	4.0	+41	MAX. P': .065	0	0	$\sqrt{40}$	1.0	0	MAX. P': .065						MAX. P':
0	0	$\sqrt{3}$	6.5	0		1	1	$\sqrt{3}$	6.5	0		0	0	$\sqrt{10}$	4.0	0							
					LOT T: 0						LOT T: +2.7	2	1	$\sqrt{3}$	6.5	-36	LOT T: -2.3						LOT T:
0	0	$\sqrt{75}$	1.0	0	LOT NO: 3045793	1	1	$\sqrt{10}$	4.0	0	LOT NO: 3014927	0	0	$\sqrt{75}$	1.0	0	LOT NO: 3077154						LOT NO:
0	0	$\sqrt{25}$	2.5	0	MAX. P': .040	0	0	$\sqrt{3}$	6.5	0	MAX. P': .065	1	3	$\sqrt{75}$	1.5	-127	MAX. P': .065						MAX. P':
0	1	$\sqrt{10}$	4.0	-41								1	1	$\sqrt{3}$	6.5	0							
					LOT T: -1.6						LOT T: 0						LOT T: -8.3						LOT T:
0	1	$\sqrt{75}$	1.0	-116	LOT NO: 3001932	3	2	$\sqrt{65}$	1.5	+68	LOT NO: 3067132						LOT NO:						LOT NO:
2	2	$\sqrt{75}$	1.5	0	MAX. P': .065	0	0	$\sqrt{25}$	2.5	0	MAX. P': .040						MAX. P':						MAX. P':
0	0	$\sqrt{3}$	6.5	0		0	0	$\sqrt{10}$	4.0	0													
					LOT T: -7.5						LOT T: +2.7						LOT T:						LOT T:
0	0	$\sqrt{100}$	.65	0	LOT NO: 3001345	0	0	$\sqrt{65}$	1.0	0	LOT NO: 3047198						LOT NO:						LOT NO:
0	0	$\sqrt{75}$	1.5	0	MAX. P': .040	4	3	$\sqrt{25}$	1.5	+110	MAX. P': .040						MAX. P':						MAX. P':
2	3	$\sqrt{75}$	2.5	-30		1	1	$\sqrt{10}$	4.0	0													
0	0	$\sqrt{10}$	4.0	0	LOT T: -1.2						LOT T: +4.4						LOT T:						LOT T:

AQL (%)	.25	.65	1.0	1.5	2.0	2.5	4.0	6.5
K	8200	1920	1000	550	360	260	130	63

C_1 = NO. OF DEFECTIVES FOUND BY REGULAR INSPECTION.
C_2 = NO. OF DEFECTIVES FOUND BY AUDIT INSPECTION.
n = SAMPLE SIZE.
M = NO. OF LOTS INSPECTED FOR ONE MONTH.
P'MAX = MAXIMUM AQL USED (DECIMAL).

CALCULATIONS

$\Sigma T = -11.1$

$M = 10$

$$\underset{\text{(for lot 3045793)}}{t} = \frac{C_1 - C_2}{\sqrt{n}} \cdot K = \frac{0-1}{\sqrt{10}} \cdot 130 = -41$$

$$\underset{\text{(for lot 3045793)}}{T} = \Sigma t \cdot p'_{\text{max.}} = -41\,(.040) = -1.6$$

$$A_R = 87 + \frac{3.3\,\Sigma T}{\sqrt{m}} = 87 + \frac{3.3(-11.1)}{\sqrt{10}} = 75$$

Fig. 6. Sampling plan record sheet

and 2. as an actual numerical rating in relation to the limits specified.

Operational Procedure

The operational procedure for the audit control of material in stock is as follows:

1. Audit inspection personnel scrutinize all areas in which inspection operations are performed. They conduct independent inspections, using the same type of inspection techniques as are normally used in the area; e.g., 100% inspection or a specific sampling plan.

2. Findings of audit inspection are reported to quality engineering on audit inspection record forms. Information originating in manufacturing areas is recorded on the form shown in Fig. 4(A); information originating in receiving inspection areas is recorded on the form shown in Fig. 4(B).

3. Quality engineering uses the data appearing on the audit inspection record forms (Fig. 4) to evaluate material passed into stock by individual inspection areas. Depending upon the area in which the data originate (manufacturing or receiving inspection), the evaluation of individual areas is computed according to the appropriate procedures described in the following paragraphs.

The procedure for the evaluation of data originating in the manufacturing area (100% inspection) is as follows:

1. The per-cent-defective of each classification of defects (solder, major, or minor) is determined as a ratio of the number of defects to a predetermined base. (Solder defects are regarded as major defects and are classified according to a particularly stringent quality level. The predetermined base is a figure which represents the complexity of the particular unit undergoing inspection. This figure is equivalent to the base figure used for the inspection of the final product.)

2. The per-cent-defective achieved is then converted into terms of the per cent-defective desired and is weighted by a constant so that when the weighted result is subtracted from 100, a quality rating figure is achieved. See Fig. 5 for a typical audit rating work sheet.

Thus, for the quality rating of solder, major, and minor characteristics, the respective formulas are as follows:

$$\text{Solder rating} = 100 - \frac{240{,}000\, d_s}{N}$$

$$\text{Major rating} = 100 - \frac{180{,}000\, d_{ma}}{N}$$

$$\text{Minor rating} = 100 - \frac{60{,}000\, d_{mi}}{N}$$

Over-all rating (entire area combining solder, major, and minor defects in their

$$\text{proper proportion}) = 100 - \frac{20{,}000}{N} \times (4d_s + 3d_{ma} + d_{mi})$$

where

N = predetermined base of the unit involved, multiplied by the number of units inspected during the month
d_s = total number of solder defects found by audit inspection personnel (for one month)
d_{ma} = total number of major defects found by audit inspection personnel (for one month)
d_{mi} = total number of minor defects found by audit inspection personnel (for one month)

The derivations of the formulas for the manufacturing areas are shown in Appendix III.

A practical example demonstrating the operational procedure for the evaluation of data originating in the manufacturing area follows.

In a particular manufacturing area, e.g., assembly of pluggable units, audit inspection personnel examined 100 pluggable units. The examination disclosed 3 solder defects, 7 major defects, and 17 minor defects. A predetermined base value of 250 was assigned to the pluggable unit assembly area. Thus, where

$d_s = 3$

$d_{ma} = 7$

$d_{mi} = 17$

$N = 250 \times 100 = 25{,}000$

then

$$\text{Solder rating} = 100 - \frac{240{,}000(3)}{25{,}000} = 71$$

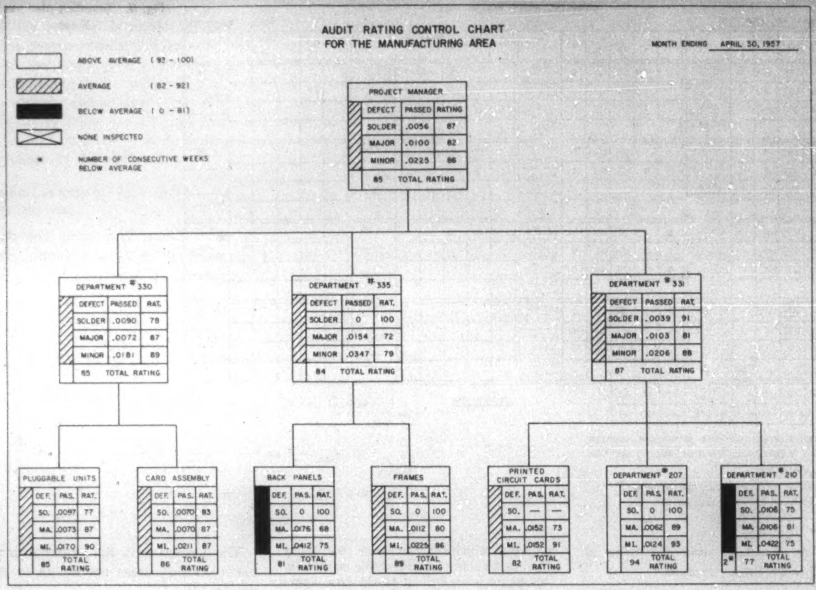
$$\text{Major rating} = 100 - \frac{180{,}000(7)}{25{,}000} = 50$$

$$\text{Minor rating} = 100 - \frac{60{,}000(17)}{25{,}000} = 59$$

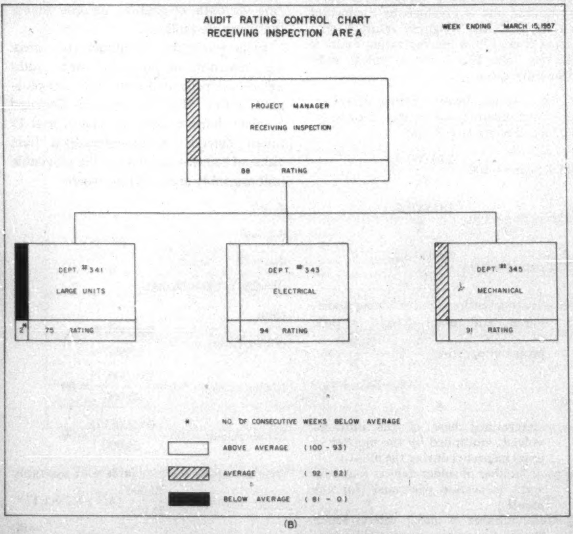
Over-all rating for pluggable unit assembly

$$\text{area} = 100 - \frac{20{,}000}{25{,}000}(4.3 + 3.7 + 1.17) = 60$$

The procedure for the evaluation of data originating in receiving inspection (sampling inspection) starts with com-


AUDIT RATING CONTROL CHART
FOR THE MANUFACTURING AREA
MONTH ENDING APRIL 30, 1957
ABOVE AVERAGE (93 - 100)
AVERAGE (82 - 92)
BELOW AVERAGE (0 - 81)
NONE INSPECTED
* NUMBER OF CONSECUTIVE WEEKS BELOW AVERAGE
PROJECT MANAGER
DEFECT PASSED RATING
SOLDER .0056 87
MAJOR .0100 82
MINOR .0225 86
85 TOTAL RATING
DEPARTMENT #330
DEFECT PASSED RAT.
SOLDER .0090 78
MAJOR .0072 87
MINOR .0181 89
85 TOTAL RATING
DEPARTMENT #335
DEFECT PASSED RAT.
SOLDER 0 100
MAJOR .0154 72
MINOR .0347 79
84 TOTAL RATING
DEPARTMENT #331
DEFECT PASSED RAT.
SOLDER .0039 91
MAJOR .0103 81
MINOR .0206 88
87 TOTAL RATING
PLUGGABLE UNITS
DEF. PAS. RAT.
SO. .0097 77
MA. .0073 87
MI. .0170 90
85 TOTAL RATING
CARD ASSEMBLY
DEF. PAS. RAT.
SO. .0070 83
MA. .0070 87
MI. .0211 87
86 TOTAL RATING
BACK PANELS
DEF. PAS. RAT.
SO. 0 100
MA. .0176 68
MI. .0412 75
81 TOTAL RATING
FRAMES
DEF. PAS. RAT.
SO. 0 100
MA. .0112 80
MI. .0225 86
89 TOTAL RATING
PRINTED CIRCUIT CARDS
DEF. PAS. RAT.
SO. — —
MA. .0152 73
MI. .0152 91
82 TOTAL RATING
DEPARTMENT #207
DEF. PAS. RAT.
SO. 0 100
MA. .0062 89
MI. .0124 93
94 TOTAL RATING
DEPARTMENT #210
DEF. PAS. RAT.
SO. .0106 75
MA. .0106 81
MI. .0422 75
2* 77 TOTAL RATING

(A)


AUDIT RATING CONTROL CHART
RECEIVING INSPECTION AREA
WEEK ENDING MARCH 15, 1957
PROJECT MANAGER
RECEIVING INSPECTION
88 RATING
DEPT. #341
LARGE UNITS
2* 75 RATING
DEPT. #343
ELECTRICAL
94 RATING
DEPT. #345
MECHANICAL
91 RATING
* NO. OF CONSECUTIVE WEEKS BELOW AVERAGE
ABOVE AVERAGE (100 - 93)
AVERAGE (92 - 82)
BELOW AVERAGE (81 - 0)

(B)

Fig. 7. Audit rating control chart

A—For manufacturing areas employing 100% inspection
B—For inspection areas employing sampling inspection

putation of the quality rating of each receiving inspection area in the following manner:

1. Determine a value (t) for each acceptance quality level (AQL) used in a particular lot:

$$t=\frac{C_1-C_2}{\sqrt{N}}\times\frac{1}{p'\sqrt{p'(1-p')}}$$

where

C_1 = defects found by regular inspection
C_2 = defects found by audit inspection
N = size of sample
p' = AQL value (decimal less than 1)

The expression $1/p'\sqrt{p'(1-p')}$ has been calculated for all AQL values used, and is interpreted as the parameter K on the sampling plan work sheet. (Derivations of formulas (for receiving inspection areas) are given in Appendix IV. For a typical sampling plan record sheet see Fig. 6. See Fig. 5 for summary data of receiving inspection areas.)

2. Determine a total rating (T) for each lot:

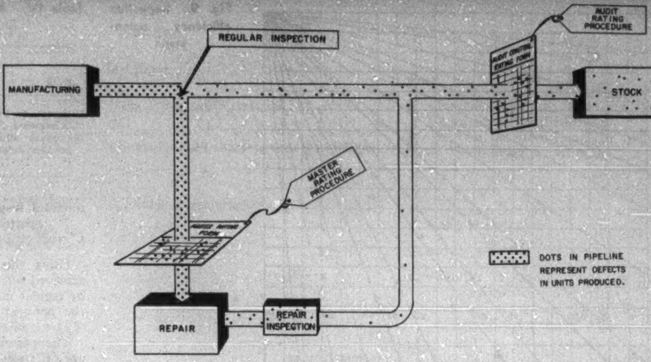
$$T=\sum_{i=1}^{j} t_i p'_{\max}$$

where

$t_i = t$ for the i^{th} AQL class
j = number of AQL classes used

3. Determine the monthly quality rating (QR)

Fig. 8. Quality assurance operation



$$QR = 87 + \frac{3.3 \sum_{e=1}^{M} T_e}{\sqrt{M}}$$

where

$T_e = T$ for the e^{th} lot

M = number of lots inspected

The constants 87 and 3.3 are chosen so that when the findings of audit inspection are equal to the findings of regular inspection, the area investigated will receive a quality rating of 87, the mid-point of the 92–82 average scale. However, if audit inspection findings indicate a 1.5 sigma increase over the number of defects found by regular inspection, the area investigated will receive a rating of 82, the lower limits of the 92–82 average scale. Similarly, a 1.5 sigma decrease in the number of defects found would cause the area to be rated 92, the upper limit of the average rating.

CHART DISPLAYS

The audit rating control chart for areas using 100% inspection (manufacturing areas) displays the per-cent-defective and quality rating for each classification of defects (solder, major, and minor) as well as the over-all rating for each manufacturing area. A typical example of the audit rating control chart for use in manufacturing areas employing 100% inspection is shown in Fig. 7(A).

The audit rating control chart for areas using sampling inspection (receiving inspection areas) displays the over-all quality rating for each inspection area. A typical example of the audit control chart for use in inspection areas employing sampling inspection is shown in Fig. 7(B).

The tell-at-a-glance information displayed on these two charts accomplishes the following results:

1. Reveals quality ratings of individual areas of inspection with respect to the ultimate requirements of those areas.
2. Calls immediate attention to areas with below-average quality ratings.
3. Specifically reveals the per-cent-defectives passed into stock; see Fig. 7(A).

Working Relation of Master Control of Manufacturing Areas and Audit Control of Material in Stock

The quality assurance procedures described in this report under master control of manufacturing areas and audit control of material in stock have a direct working relation with each other as shown in the flow diagram of Fig. 8.

Upon completion of the manufacturing operation, all units are inspected 100% by regular inspection. Any defect which is discovered by regular inspection must be repaired by manufacturing before the unit is considered acceptable as a finished product ready for audit inspection. The maximum per-cent-defective which regular inspection could find for a manufacturing area (if the end product is to fall within tolerance limits) was determined on the assumption that regular inspection would maintain a minimum efficiency of 82%. This gives a method whereby regular inspection can report on the quality or work accomplished by manufacuring and, at the same time, control the quality of the components which make up the end product. This function is described as the master control of manufacturing areas.

The next step is to check the quality of material as it enters into stock in preparation for assembly into the final system. This operation is described as the audit control of material in stock. Here audit inspection determines whether the final product meets requirements and rates the product with respect to these requirements. This rating is a numerical evaluation which reveals the quality of the product after it has been produced by manufacturing and examined by regular inspection.

Up to this point, therefore, the master control plan shows the quality of the product of the manufacturing operation, while the audit control plan gives vital information regarding the quality of material which goes into stock.

There remains now the need for determining the actual efficiency with which regular inspection is able to detect poor quality. This inspection efficiency is determined in the following manner:

$$\text{Inspection efficiency} = \frac{P_t}{P_t + P_a}$$

where

P_t = per-cent-defective found by regular inspection

P_a = per-cent-defective overlooked by regular inspection

but (from master rating plan):

$$M_r = 100 - \frac{KP_t}{C'} \text{ and } P_t = \frac{(100 - M_r)C'}{K}$$

where

M_r = master rating

K = constant = 13

C' = per-cent-defective in manufacturing which would give a rating of 87

and (from audit rating plan):

$$A_r = 100 - \frac{KP_a}{P'}$$

$$P_a = \frac{(100 - A_r)P'}{K}$$

where

A_r = audit rating

K = constant = 13

P' = per-cent-defective in the final product which would give an average rating of 87

Interpreting P_t and P_a in terms of the master rating and audit rating:

$$\frac{P_t}{P_t + P_a} = \frac{\text{Inspection}}{\text{Efficiency}} = \frac{100 - M_r}{122 - 0.22A_r - M_r}$$

For ease of interpretation, a nomo-

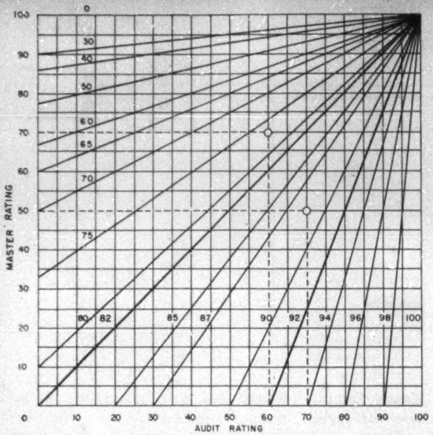


Fig. 9. Inspection efficiency nomograph

graph (Fig. 9) was constructed based on this formula. To find the efficiency of any regular inspection area it is only necessary to find the point of intersection of the master rating with the audit rating.

As with both the master control and audit control plans, ratings appear as follows:

0–82 = below average inspection efficiency
82–92 = average inspection efficiency
92–100 = above average efficiency

Conclusions

Since the adoption of this quality assurance procedure by IBM Kingston, the resulting tell-at-a-glance charts have enabled management to review the entire manufacturing process with significant improvement in ease of locating and correcting workmanship errors.

Appendix I. Derivation of Maximum Per-Cent-Defective Allowance to Manufacturing Found by Regular Inspection

For the ultimate per-cent-defectives allowable in the end product for each classification of defects, see Table II.

The fraction-defective which regular inspection finds (P_t) and the fraction-defective which is overlooked by regular inspection (P_a) are the total fraction-defective produced by manufacturing (P_m). Therefore, $P_m = P_t + P_a$.

Assume a minimum inspection efficiency of 82%. Then

$$\frac{P_t}{P_m} = 0.82$$

Therefore

$$P_m = \frac{P_t}{0.82}$$

Substitute the value of P_m into the original equation, and

$$P_t = \frac{0.82}{0.18} P_a$$

Substituting for P_a the per-cent-defective allowable in the end product for solder, major, and minor characteristics, respectively, determine the maximum per-cent-defective which regular inspection should find (at an inspection efficiency of 82%) if the end product is to meet the ultimate quality requirements. These maximum per-cent-defectives allowable (P_t) are found in Table IV.

Appendix II. Derivation of Master Rating of Manufacturing Area

Conditions or assumptions:

1. The lowest acceptable quality rating shall be 82.
2. A master rating of 82 will be given when regular inspection finds a per-cent-defective equal to the maximum per-cent-defective which regular inspection should find (P_t). See Table III.
3. A master rating of 100 is given when regular inspection finds no defects.

Formula:

$$M_r = 100 - \frac{K\,100\,\frac{d}{n}}{C'}$$

where

M_r = master rating
K = constant = 13
d = number of defects found by regular inspection

Table IV. Maximum Allowable Per-Cent-Defective

	Solder, %	Major, %	Minor, %
Ultimate quality requirements, P_a	0.0075	0.01	0.03
Maximum allowable per-cent-defective, P_t	0.034	0.045	0.136

n = total inspected; number of units inspected times the complexity base
C' = per-cent-defective considered average

Using this formula, a rating of 87 is achieved when the per-cent-defective found by regular inspection ($100\,d/n$) is equal to the per-cent-defective considered average (C').

The per-cent-defective considered average (C') was determined as that value for which a rating of 82 would be achieved when the per-cent-defective found by regular inspection ($100\,d/n$) is equal to the maximum allowable per-cent-defective (P_t).

Formula:

$$82 = 100 - \frac{13 \times P_t}{C'}$$

The C' for each P_t is found in Table V.

Table V. C′ for Each P_t

Defect Classification	P_t, %	C', %
Solder	0.034	0.0246
Major	0.045	0.0325
Minor	0.136	0.098

Using each C' as a base, the rating formulas for each type of characteristic are as follows:

$$M_r\,(\text{solder characteristics}) = 100 - \frac{13 \times 100\,\frac{d}{n}}{C'}$$

$$= 100 - \frac{1{,}300\,\frac{d}{n}}{0.0246}$$

$$= 100 - 53{,}000\,\frac{d}{n}$$

(approximately)

Therefore, in order to find the master rating for solder characteristics, it is necessary simply to multiply the per-cent-defective found by 530 and subtract this quantity from 100.

Likewise, for major and minor characteristics:

$$M_r\,(\text{major characteristics}) = 100 - 40{,}000\,\frac{d}{n}$$

$$M_r\,(\text{minor characteristics}) = 100 - 13{,}200\,\frac{d}{n}$$

(approximately)

The over-all master rating can be calculated by averaging the three values (M_r)

Vol. 7
Pt. 1
1959

Table VI. P' for Each P_e

Defect Classification	P_e, %	P', %
Solder	0.0075	0.0054
Major	0.01	0.0072
Minor	0.03	0.0216

found separately since these values have already been properly weighted.

The over-all master rating can also be calculated by combining the individual formulas and obtaining:

Over-all master rating (for any area)

$$=100-\frac{43{,}000}{n}(4d_s+3d_{ma}+d_{mi})$$

Appendix III. Derivation of Audit Rating of Manufacturing Area

Conditions or assumptions:

1. Audit inspection finds all defects missed by regular inspection.
2. The lowest rating indicating acceptable quality is 82.
3. The 82 rating is given when audit inspection finds a per-cent-defective equal to the maximum per-cent-defective allowable in the final product; i.e.,

Solder, 0.0075% defective (maximum)
Major, 0.01% defective (maximum)
Minor, 0.03% defective (maximum)

4. The 100 rating is given when audit inspections find no defects.

Formula:

$$A_r=100-\frac{K\,100\,\frac{d}{n}}{P'}\text{ for each classification of characteristics}$$

where

A_r = audit rating
K = constant = 13
d = number of defects found by audit inspection
n = sample size (number of units inspected times the complexity base)
P' = per-cent-defective considered average

Using this formula, a rating of 87 is achieved when the per-cent-defective found by audit ($100\,d/n$) is equal to the per-cent-defective considered average (P').

The per-cent-defective considered average (P') was determined as that value for which a rating of 82 would be achieved when the per-cent-defective found by audit inspection ($100\,d/n$) is equal to the maximum allowable per-cent-defective (P_e).

Formula:

$$82=100-\frac{13\cdot P_e}{P'}$$

The P' for each P_e is found in Table VI.

Using each P' as a base, the rating formula for each characteristic is as follows:

$$A_r\text{ (solder characteristics)}=100-\frac{13\times 100\,\frac{d}{n}}{P'}$$

$$=100-\frac{1{,}300\,\frac{d}{n}}{0.0054}$$

$$=100-240{,}000\,\frac{d}{n}$$

(approximately)

Therefore, in order to find the audit rating for solder characteristics, it is necessary simply to multiply the per-cent-defective found by 2,400 and subtract this quantity from 100.

Formulas for major and minor classifications of defects can be found by this procedure and are as follows:

$$A_r\text{ (major characteristics)}=100-180{,}000\,\frac{d}{n}$$

(approximately)

$$A_r\text{ (minor characteristics)}=100-60{,}000\,\frac{d}{n}$$

(approximately)

The over-all audit rating can be found by averaging the three separately found A_r values, and can also be calculated by combining the individual formulas:

$$\text{Over-all audit rating}=100-\frac{20{,}000}{n}(4d_s+3d_{ma}+d_{mi})$$

Appendix IV. Derivation of Basic Formulas for Receiving Inspection Found by Sampling Inspection Areas

Let the standard deviation, σ, be computed for that fraction defective which is considered acceptable:

$$\sigma_p=\frac{\sqrt{p'(1-p')}}{n}$$

where

p' = AQL value (fraction defective)
n = sample size

Let t be a measure of the difference between regular inspection results and audit inspection results in terms of $\sigma_{p'}$ for each AQL used, then:

$$t=\frac{p_1-p_2}{\sigma_{p'}}\times\frac{1}{p'},\text{ for the }i^{\text{th}}\text{ AQL}$$

where

p_1 = fraction defection found by regular inspection for the i^{th} AQL
p_2 = fraction defective found by audit inspection for the i^{th} AQL

$$t=\frac{\frac{C_1}{n}-\frac{C_2}{n}\sqrt{n}}{p'\sqrt{p'(1-p')}}$$

Therefore, where

C_1 = number of defects found by regular inspection for the i^{th} AQL
C_2 = number of defects found by audit inspection for the i^{th} AQL

or

$$t=\frac{C_1-C_2}{p'\sqrt{n}}\times\frac{1}{\sqrt{p'(1-p')}}$$

let

$$\frac{1}{\sqrt{p'(1-p')}}=K$$

K is calculated for each AQL used. The table of K values appears on the sampling plan work sheet of Fig. 6.

The t for the i^{th} AQL (t_i) is simply:

$$t_i=\frac{C_1-C_2}{\sqrt{n}}\times K_i$$

A T is calculated for the e^{th} lot:

$$T_e=\sum_{i=1}^{J}t_i\times P'_{\max}$$

where

J = number of AQL's involved in the inspection of the lot
$p'_{\max}$ = maximum AQL (in fraction defective) involved in the inspection of the lot

The monthly audit rating (A_r) is calculated for each inspection area:

$$A_r=87+\frac{3.3\sum_{e=1}^{M}T_e}{\sqrt{M}}$$

where

87 = arbitrary constant chosen such that when audit inspection finds a number of defects equal to the number of defects found by regular inspection, an inspection area will receive a rating of 87
3.3 = an arbitrary constant chosen such that when audit inspection finds a number of defects equivalent to 1.5 greater than regular inspection, an inspection area will receive a rating of 82
M = number of lots involved in the monthly rating for an area

Application of Telegraph Techniques in Data Transmission

A. BOGGS
NONMEMBER AIEE

J. E. BOUGHTWOOD
ASSOCIATE MEMBER AIEE

Synopsis: The impact of modernization and mechanization in business has brought drastic changes in methods and instrumentation to handle large volumes of data at high speeds. Such high-speed handling has much in common with conventional telegraphy, and telegraph techniques in principle apply to data transmission systems.

THE DEVELOPMENT of a transmission system, or systems, to handle data on a nation-wide basis must be based on a knowledge of the various types of service that may be required. It may not be possible to forsee all potential requirements at the present state of the art but some, at least, are clearly indicated. Two of these are:

1. Point-to-point service, intracity or intercity, either private wire or shared facilities.
2. Multipoint service, usually private wire, such as ticket reservation or the bookkeeping applications of large corporations.

Point-to-point service for private-wire patrons poses no serious problems and can be provided in a straightforward manner once the speed of operation has been determined. Shared facilities, such as the use of an intercity circuit by several patrons in each city, introduces some complications. Manual or automatic switching between local patron facilities and the intercity circuit must be provided. For multipoint service, where groups of patrons in different cities must be able to interchange data information, still more complicated mechanisms are required.

Carrier Equipment

For a number of years the Western Union Telegraph Company (WU) has supplied data-handling circuits engineered to fit in with its standard telegraph plant. These circuits have operated at speeds which would allow them to be assigned to narrow-band 75-baud carrier channels normally used for teleprinter service. Our planning in respect to higher-speed comprehensive data systems has been influenced by our experience with these lower-speed circuits and by consideration of transmission rates likely to represent the largest immediate use. At the same time compatibility with our present carrier telegraph plant has been preserved.

It is standard practice in WU to divide a voice band into two subbands, each capable of carrying 10 telegraph channels spaced at 150-cycle intervals. By assigning the frequency space occupied by two adjacent telegraph channels to a single data channel, a spectrum width adequate for 180 bauds is made available. This is the bit rate required by the 11-cards-per-minute IBM (International Business Machines Corporation) transceiver and represents the most common demand of commercial users of data equipment at the present time. Equipment for two such channels, one centered at 1,050 cps (cycles per second) and the other at 1,350 cps, is now available.

For rates higher than 180 bauds the entire subband spectrum, 300 cps to 1,800 cps, is utilized. The carrier equipment developed for this service is capable of operating at a maximum rate of 1,000 bauds.

At present, provision of equipment for rates higher than 1,000 bauds is not contemplated. However, circumstances are envisioned wherein it may be preferable to abandon the subband and position the 1,000 baud system in an optimum region of a whole voice band. This can be done easily with the present equipment.

Terminal Handling

The varied types of service which may be required, particularly the interconnection of a multiplicity of drops with carrier terminal legs on a half-duplex basis at a multipoint hubbing center, and the desirability of utilizing standardized equipment throughout, led to the adoption of d-c transmission on patrons' loops rather than the use of extended carrier legs. This d-c loop equipment has been engineered to accommodate a transmission speed of 1,000 bauds over balanced pairs with a steady-state loop current of 3 to 5 ma (milliamperes). Only simple *RC* equalization networks are required at this speed.

Interconnection

An important consideration in data handling by a common carrier involves the actual interconnection of the data processing equipment at the patron's location with the common carrier's equipment also at the patron's location. It is quite obvious that standards must be set if confusion is to be avoided. WU follows the standards proposed in reference 1 for the particular type of service furnished to the patron. In addition, facilities for the alternate use of the data circuit for teleprinter communication, and means for checking the common carrier transmission equipment at the patron's office, are also provided.

Data Systems

Fig. 1 illustrates one version of a point-to-point system furnishing one 180-baud data channel with New York and Chicago terminals, and one 180-baud channel via Chicago with New York and Omaha terminals. Two data channels, *A* and *B*, centered at 1,050 and 1,350 cps respectively, and six regular telegraph channels at 375, 525, 675, 825, 1,575, and 1,725 cps occupy one subband between New York and Chicago. The legs of carrier transceiver *A* are patched to the legs of transceiver *C* to establish the New York-to-Omaha connection. The Chicago-to-Omaha subband also carries eight telegraph channels, one channel at 975 and one at 1,125 replacing a 1,050-cps data channel not used in this section.

The legs of carrier transceiver *B* are patched to the legs of loop transceiver *D*. This transceiver converts the grounded polar signals of the carrier transceiver leg to a form suitable for transmission to the patron's premises over a balanced pair. Loop transceiver *E* on the patron's premises converts the received loop signals back to grounded polar form.

At the patron's office the data circuit is used to operate an 11-cards-per-minute IBM transceiver. For this particular application switching panel *F* is required to match the 10-ma legs of the loop transceiver to the higher impedance legs of the IBM set. The switching panel also permits use of the data channel for printer communication and provides means for testing the operation of the loop transceiver. Data, printer, or test function

---

Paper 59-107, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 27, 1958; made available for printing December 3, 1958.

A. BOGGS and J. E. BOUGHTWOOD are with the Western Union Telegraph Company, New York, N. Y.

Vol
Pt.
19

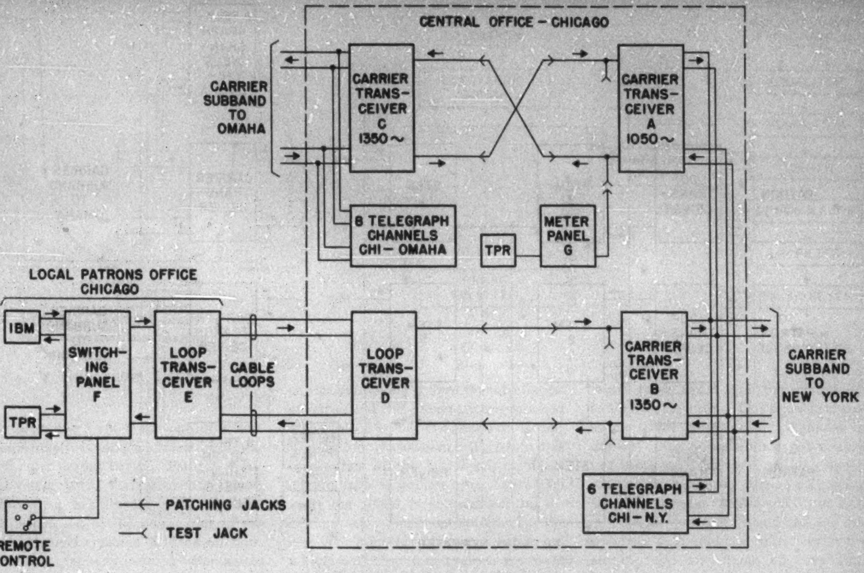

CENTRAL OFFICE – CHICAGO
CARRIER SUBBAND TO OMAHA
CARRIER TRANS-CEIVER C 1350 ~
CARRIER TRANS-CEIVER A 1050 ~
8 TELEGRAPH CHANNELS CHI – OMAHA
TPR
METER PANEL G
LOCAL PATRONS OFFICE CHICAGO
IBM
SWITCH-ING PANEL F
LOOP TRANS-CEIVER E
CABLE LOOPS
LOOP TRANS-CEIVER D
CARRIER TRANS-CEIVER B 1350 ~
CARRIER SUBBAND TO NEW YORK
TPR
6 TELEGRAPH CHANNELS CHI – N.Y.
PATCHING JACKS
TEST JACK
REMOTE CONTROL

Fig. 1. Point-to-point service, typical installation

may be selected by remote control from an operations center.

At the WU central office, meter panel *G* and associated printer form a common piece of test gear which may be corded into any data circuit. It permits printer communication either over the carrier link or over the patron's loop and also provides for metering of signals in the transceiver legs without interrupting the circuit.

Fig. 2 illustrates equipment arrangement at Chicago on a multipoint system which includes New York, Chicago, and Omaha. It is to be understood that similar equipment appears at New York and Omaha, and also that the system is not limited to three hubbing points. Three are shown here only for convenience.

Chicago patrons are connected into the network through the network repeater. Functions of other equipment shown are the same as was outlined in the point-to-point system of Fig. 1. With the arrangement shown operation is half duplex, whereby any one station communicates with all others in the network. A program-type operation, in which each station uses the circuit only on invitation to query a central computer and receive a reply, is in development.

The factor limiting transmission speed in either system is the 180-baud carrier channel. By using a full subband with delay correction and regenerative repeaters where necessary, operation at 1,000 bauds is possible throughout the entire network.

Data Transmission Equipment

All of the circuit equipments indicated on Figs. 1 and 2 have been developed and are now in limited production. The carrier transceiver for 1,000-baud operation has also been developed and prototypes tested on actual circuits. It should be of interest to cover briefly the theory of the most important items.

180-Baud Data Carrier Transceiver

This component shown in Fig. 3 is a directly modulated frequency-shift transceiver operable on a carrier frequency of 1,050 or 1,350 cps. The unit is completely transistorized and conventional frequency-shift techniques are used throughout. Marking frequency is 80 cps below and spacing frequency 80 cps above channel mid-frequency to produce a 160-cycle shift. The impedance of the sending leg is 620 ohms to ground and the impedance of the receiving leg 620 ohms to 12.5-volt polar battery.

The band-pass channel filters are conventional in design except that some departure from usual practice is introduced in the sending filter to reduce delay distortion. Adequate suppression of higher order sidebands of the wide-band frequency-shift keying requires this filter to have a pass band narrower than that required in the receiving filter. Insertion-loss and envelope-delay characteristics of the sending filter of the 1,350-cps channel are shown on Fig. 4. Both sending and receiving filters are designed to parallel with the standard 600-ohm telegraph channel filters.

Following the receiving filter, a 2-stage push-pull limiter provides 35 db (decibels) of limiting and the required power output for driving the discriminator. Two series-tuned circuits are used in the discriminator one tuned below the marking frequency and the other above the spacing frequency. The currents from the two circuits are individually rectified and differentially added to produce a polar signal output current. The signal shaping network discriminates against frequencies above 90 cps and shapes the signal for minimum distortion. Polar current from the filter drives a 2-stage directly coupled,

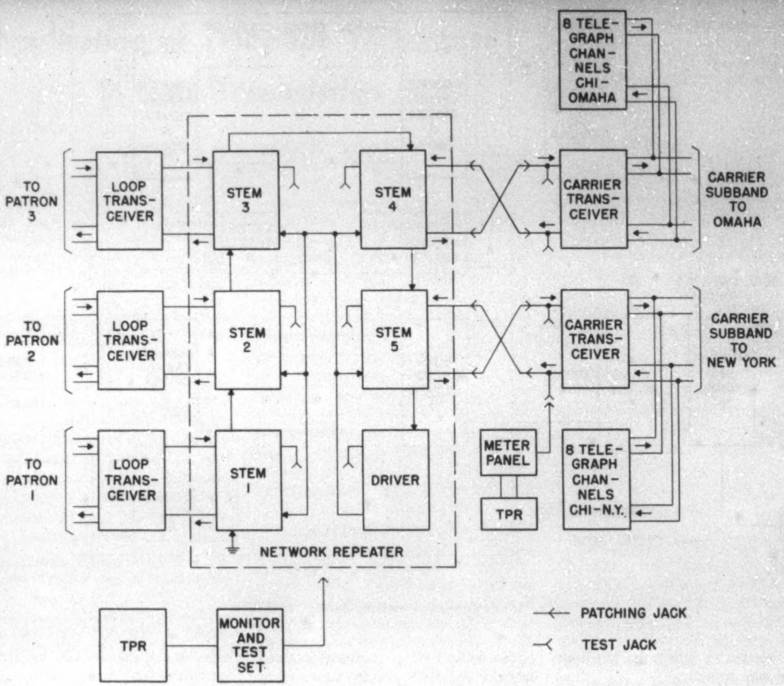

8 TELEGRAPH CHANNELS CHI-OMAHA
TO PATRON 3
LOOP TRANSCEIVER
STEM 3
STEM 4
CARRIER TRANSCEIVER
CARRIER SUBBAND TO OMAHA
TO PATRON 2
LOOP TRANSCEIVER
STEM 2
STEM 5
CARRIER TRANSCEIVER
CARRIER SUBBAND TO NEW YORK
TO PATRON 1
LOOP TRANSCEIVER
STEM 1
DRIVER
METER PANEL
TPR
8 TELEGRAPH CHANNELS CHI-N.Y.
NETWORK REPEATER
TPR
MONITOR AND TEST SET
PATCHING JACK
TEST JACK

Fig. 2 (above). Multipoint service; typical installation

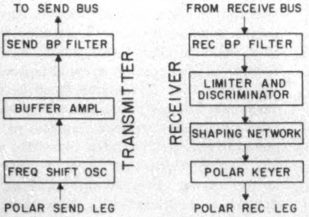

TO SEND BUS
SEND BP FILTER
BUFFER AMPL
FREQ SHIFT OSC
POLAR SEND LEG
TRANSMITTER
RECEIVER
FROM RECEIVE BUS
REC BP FILTER
LIMITER AND DISCRIMINATOR
SHAPING NETWORK
POLAR KEYER
POLAR REC LEG

Fig. 3 (left). The 180-baud data carrier transceiver

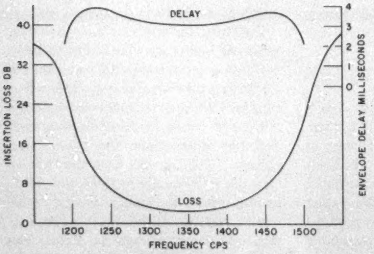


Fig. 4. The 180-baud data carrier transceiver, sending filter characteristics

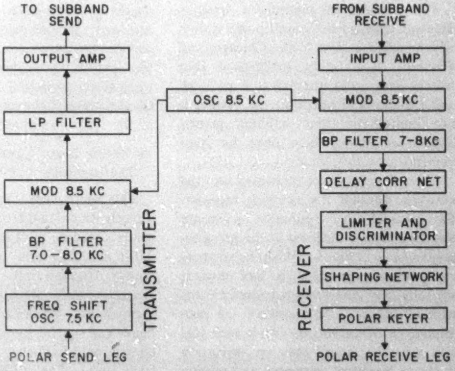

TO SUBBAND SEND
OUTPUT AMP
LP FILTER
MOD 8.5 KC
BP FILTER 7.0–8.0 KC
FREQ SHIFT OSC 7.5 KC
POLAR SEND LEG
TRANSMITTER
OSC 8.5 KC
FROM SUBBAND RECEIVE
INPUT AMP
MOD 8.5 KC
BP FILTER 7–8KC
DELAY CORR NET
LIMITER AND DISCRIMINATOR
SHAPING NETWORK
POLAR KEYER
POLAR RECEIVE LEG
RECEIVER

Fig. 5. The 1,000-baud data carrier transceiver

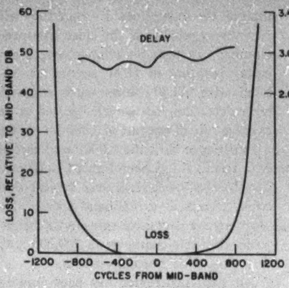


Fig. 6 (above). The 1,000-baud data carrier transceiver transmission characteristics

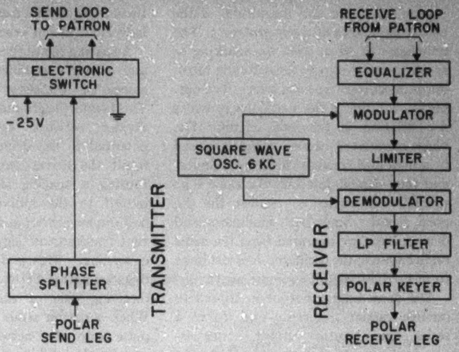


Fig. 7 (right). Data loop transceiver

complementary transistor output circuit which functions as a post-detection limiter and polar output keyer. The polar output voltage is 12.5 volts through 620 ohms. Thus when the receiving leg of one channel is patched to the sending leg of another channel a current of 10 ma will flow. The unit contains rectifier supplies to provide minus 25 volts and plus and minus 12.5 volts from 115 volts a-c. When operating over a normal voice band, maximum distortion introduced by the equipment is 5% characteristic and 2% fortuitous.

1,000-Band Data Carrier Transceiver

The transceiver shown in Fig. 5 has been developed to utilize a full subband for speeds up to 1,000 bauds. Again frequency-shift operation is used with a center frequency of 1,000 cycles and a deviation of 400 cycles, the spacing and marking frequencies being 600 and 1,400 cycles respectively. To transmit at such a high modulation-frequency-to-carrier-frequency ratio over a medium which does not rigidly preserve this ratio precludes the use of direct modulation. Also, demodulation of a frequency-shift signal uses only the crossover information of the received carrier signal and so provides insufficient samples to define the modulation envelope completely.

These difficulties are avoided by using a double modulation method. The frequency-shift signal is generated at 7.5 kc and with a deviation of ±400 cycles, and passed through a band-pass filter to restrict the sideband energy to a 1,400-cycle band. The carrier signal is then modulated with an 8.5-kv carrier and the lower side band selected for transmission over the subband. The 7.5-kc carrier adequately defines the modulation envelope and this information is retained when the 7.5-kc carrier signal is translated to 1 kc for transmission. To retain this information at the 1-kc level, nonlinear circuits such as limiters must be avoided until the signal is retranslated to the 7.5-kc level. This is accomplished at the receiver by another 8.5-kc modulator, and a band-pass filter which selects the lower sideband in the 6.8- to 8.2-kc range so that the signal is restored to its original high-frequency state. At this point a delay-correction network compensates for the delay distortion introduced by the sending and receiving circuits of the transceiver. The transmission characteristics with the corrective network added are shown on Fig. 6. The signal is then limited and demodulated in the conventional manner and delivered to the receiving leg. Again both sending and receiving legs operate with 10-ma polar signals.

On a back-to-back basis the maximum characteristic distortion at speeds up to 1,000 bauds is 2% to 3%. In normal operation external networks correct the delay distortion introduced by the subband modulating equipment and carrier-system voice bands to limit the characteristic distortion to about twice the back-to-back value.

Data Loop Transceiver

For reasons previously mentioned, transmission between the central telegraph office and the patron is on a d-c or video basis and separate sending and receiving loops are employed. Transmission is accomplished by data loop transceivers located both at the central office and the patron's premises.

A block diagram of the loop transceiver is shown on Fig. 7. The transmitter phase splitter converts the 10-ma ground-return polar signal from the sending leg to a push-pull form. This signal operates an electronic switch to apply ground to one side of the loop and a negative voltage to the opposite side for a marking signal, or to reverse these conditions for a spacing signal. Since the distant end of the loop is terminated free of ground and the impedances to ground and to battery at the sending end are equal, effective suppression of extraneous interference is obtained. For this reason steady-state loop currents of 5 ma or less permit distortion-free operation at 1,000 bauds over cables having loop resistances of 5,000 ohms or more.

The receiving portion of the transceiver accepts signals from the loop and applies them to a balanced half-section *L/C* filter to reject components above 500 cycles. The signals are then equalized for amplitude and phase in a simple balanced *RC* equalizer. The equalizer is adjusted for equal output at 25-cycle and 500-cycle dotting signals. A double balanced modulator driven by a 6-kc square-wave carrier converts the d-c push-pull signal from the equalizer to a 180-degree phase-modulated carrier signal. One phase represents a marking signal and the opposite phase a spacing signal. Bias control is affected by ad-

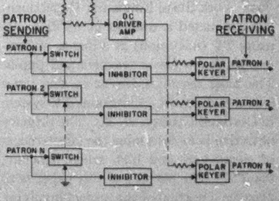


Fig. 8. Data network repeater

justing the sense and magnitude of the carrier unbalance of the modulator. The unbalance in phase with the marking or spacing carrier signal, respectively, introduces marking or spacing bias. The output transformer of the modulator provides d-c isolation of the loop circuit. The phase-modulated carrier signal is raised to a high and constant level by a limiter and then demodulated by the same 6-kc square-wave carrier to recover the d-c polar signal. Since both modulator and demodulator are operated from the same carrier source, the oscillator does not have to meet rigid stability requirements.

The demodulator output is filtered to remove carrier transients and drives a series complementary output limiter circuit to deliver 10-ma polar signals to the receiving leg.

Network Repeater

The transistorized network repeater is shown in Fig. 8 and differs somewhat from its relay equivalent currently in use for telegraph networks. Each patron's local sending loop is terminated in a p-n-p transistor switch. All these switches are series connected between ground and a d-c driver amplifier. When all the patrons' sending loops are on marking (negative battery) all the transistor switches are closed and positive signal is applied by a resistor-divider network to the driver amplifier. Under this condition the driver amplifier applies a negative signal to all the polar keyers causing them to transmit a marking signal to their individual patron's receiving loops.

A spacing signal (positive battery) from any patron's sending loop causes the associated transistor switch to open. This breaks the ground connection and the divider network applies a negative potential to the driver amplifier. As a result the driver output goes positive, causing a spacing signal to be transmitted to the individual patrons. In half-duplex service it is necessary to prevent transmission back to the patron who is sending. This is done by inhibit circuits connected between the sending and receiving side of each network stem. When a patron sends a positive spacing pulse into the network repeater, the associated inhibit circuit clamps the polar keyer in the marking condition by overriding the signal from the d-c driver amplifier. The polar keyers for the other patrons are not affected because of the isolating resistors between the driver amplifier and each keyer.

Both sending and receiving legs operate on a 10-ma polar basis. As many as 20 circuits may be hubbed with no difficulty. As a result of the transistor circuitry and because only local transmission circuits are involved, operating speeds are in no way limited by the network repeater.

Conclusions

Although our experience with these higher-speed data circuits has been limited, the indications are that they will be fully as reliable as the older low-speed circuits. In tests on average length circuits operating in IBM service at 180 bauds with full 80-column cards and with error detection, cards were rejected at an average rate of one out of 3,300 for errors.

Experience with the 1,000-baud carrier data terminal has been limited to laboratory back-to-back tests over typical subband circuits at a 600-baud speed. Indications are that error rates in the neighborhood of one character in 300,000 may be expected on normal circuits.

Our field experience has been limited to point-to-point service but all necessary transmission equipments for multipoint network service have been developed and are currently available. Multipoint system circuit layouts to meet the specific requirements of several potential users of this equipment are already in the advanced planning stage and initial installations for field trials will soon be made.

With this ground work completed we believe that the foundations have been laid for a comprehensive data handling system readily adaptable to a variety of high-speed services. This system can be realized with a minimum reorganization of present operating practices and a maximum utilization of existing facilities.

Reference

1. Interconnection of Data Terminal Equipment with a Data Communications Channel. *Standards Proposal No. 576*, Electronic Industries Association, New York, N. Y., May 23, 1958.

A Transmission Design for Closed Circuit Educational Television

W. L. WALL
NONMEMBER AIEE

OVER the past few years the New York State Education Department has conducted a number of experiments using TV (television) as a means of communication between teachers and pupils. The most recent endeavor in this series is a closed circuit educational TV system in Cortland County, N. Y. The primary objective is to determine the effectiveness of elementary and high school teaching via TV. The purpose of this paper is to describe a portion of the project involving the New York Telephone Company, namely, the transmission of video and audio signals from the point of origination to the schools where the information is presented to the pupils.

This educational TV system is primarily experimental in nature. It is expected that data obtained from the operation will serve as a guide in determining the future course of teaching by TV in the state of New York. An effort has been made to provide a flexible tool which will permit various types of educational experiments and thus determine which method will best convey the subject matter from teacher to pupil.

From the transmission engineer's standpoint, Cortland County appears to be an ideal location for this project. Within the city of Cortland itself are six elementary schools and one large high school, all situated in a relatively small area. Furthermore, within a radius of about 15 miles, there are numerous central school districts which will benefit by interconnection with the Cortland TV system. Two such locations, Virgil and Truxton, N. Y., are included in the initial layout shown in Fig. 1. It is interesting to note that this will permit these outlying schools

Paper 59-200, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 17, 1958; made available for printing January 30, 1959.

W. L. Wall is with the New York Telephone Company, Syracuse, N. Y.

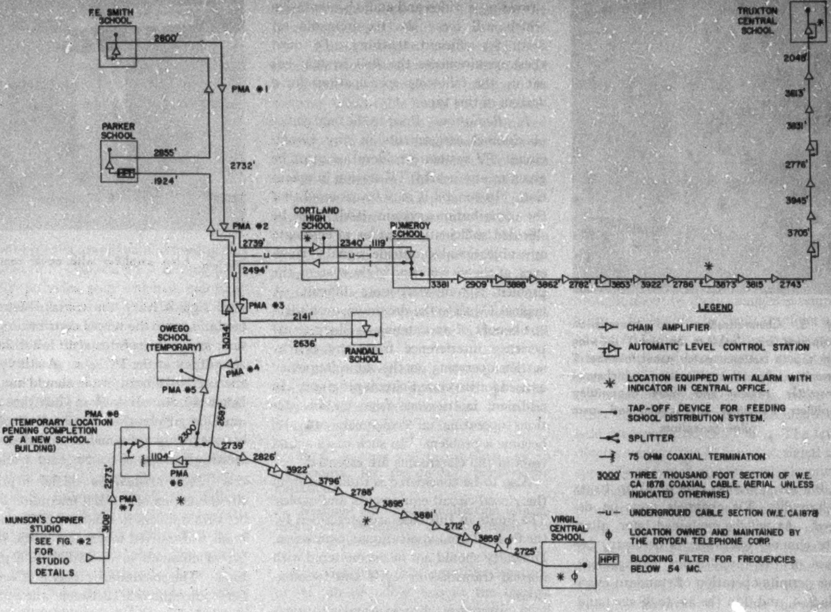

F.E. SMITH SCHOOL
2800'
PMA #1
PARKER SCHOOL
2855'
HPF
1924'
2732'
PMA #2
CORTLAND HIGH SCHOOL
2739'
2494'
2340'
1119'
POMEROY SCHOOL
3381'
2909'
3886'
3862'
2782'
3853'
3922'
2786'
3873'
3815'
2743'
3705'
3945'
2776'
3831'
3613'
2046'
TRUXTON CENTRAL SCHOOL
PMA #3
2141'
2636'
RANDALL SCHOOL
OWEGO SCHOOL (TEMPORARY)
3034'
PMA #4
PMA #5
2687'
2300'
PMA #8 (TEMPORARY LOCATION PENDING COMPLETION OF A NEW SCHOOL BUILDING)
1104'
PMA #6
2273'
PMA #7
3108'
MUNSON'S CORNER STUDIO
SEE FIG. #2 FOR STUDIO DETAILS
2739'
2826'
3922'
3796'
2786'
3895'
3881'
2712'
3859'
2725'
VIRGIL CENTRAL SCHOOL
LEGEND
CHAIN AMPLIFIER
AUTOMATIC LEVEL CONTROL STATION
LOCATION EQUIPPED WITH ALARM WITH INDICATOR IN CENTRAL OFFICE.
TAP-OFF DEVICE FOR FEEDING SCHOOL DISTRIBUTION SYSTEM.
SPLITTER
75 OHM COAXIAL TERMINATION
3000' THREE THOUSAND FOOT SECTION OF W.E. CA 1878 COAXIAL CABLE. (AERIAL UNLESS INDICATED OTHERWISE)
—u— UNDERGROUND CABLE SECTION (W.E. CA1878)
ϕ LOCATION OWNED AND MAINTAINED BY THE DRYDEN TELEPHONE CORP.
HPF BLOCKING FILTER FOR FREQUENCIES BELOW 54 MC.

Fig. 1. Cortland County educational TV system

to add subjects to their curriculums which would otherwise have been impossible.

General Description

Separate video and audio signals are supplied by the customer to the telephone company patch panel; see Fig. 2. These signals modulate miniature TV transmitters and are converted to a form which is identical to that transmitted from a conventional TV station. The various TV signals plus a 77-mc (megacycle) pilot signal, which is used for regulating purposes, are combined and transmitted out over a coaxial cable distribution system to the schools where the programs are to be viewed. Signal levels along the cable are maintained within prescribed limits through the use of line amplifiers spaced on an average of 3,000 feet. Where feasible, these amplifiers are pole-mounted. However, to provide additional flexibility, an amplifier is located at most schools.

At each school location a tap-off feeding the building distribution system is provided. The distribution system within the school as well as the classroom TV receivers are the customers' responsibility. Generally, the level of the signal at the school tap-off is adequate to handle a limited number of classrooms. As the number of classrooms equipped for viewing increases, a point will be reached where distribution amplifiers will be required in order to maintain sufficient signal level at the TV sets.

It will be seen in Fig. 1 that the portion of the system within the city of Cortland forms a modified loop. With this loop or "round-robin" configuration it is possible for any of the Cortland schools to be set up as a studio and to feed programs to the other schools. Filters permit the simultaneous transmission of this program as well as any programs originating in the regular studios. The extensions to the Truxton and Virgil schools are not included in the round-robin, and therefore these locations are not available as program origination points.

In this installation only two channels were placed in service initially; ultimately a total of five channels is anticipated. The basic distribution system provides a bandwidth which extends from about 20

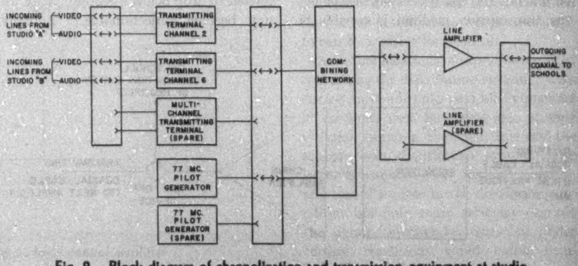

INCOMING LINES FROM STUDIO "A"
VIDEO
AUDIO
INCOMING LINES FROM STUDIO "B"
VIDEO
AUDIO
TRANSMITTING TERMINAL CHANNEL 2
TRANSMITTING TERMINAL CHANNEL 6
MULTI-CHANNEL TRANSMITTING TERMINAL (SPARE)
77 MC. PILOT GENERATOR
77 MC. PILOT GENERATOR (SPARE)
COM-BINING NETWORK
LINE AMPLIFIER
LINE AMPLIFIER (SPARE)
OUTGOING COAXIAL TO SCHOOLS

Fig. 2. Block diagram of channelization and transmission equipment at studio

Fig. 3. Channelization and transmission equipment bay (cabinet on right) showing from top to bottom: patch panel, channel 2 transmitter, channel 6 transmitter, multichannel transmitter, regular and spare transmitting amplifiers, TV tuner, and regular and spare pilot generators

to 100 mc. Within this band up to six high-quality TV channels may be derived. As will be explained later, alternate channel operation is necessary because of the round-robin arrangement. This permits operation of standard channels 2, 4, and 6 in the 54- to 88-mc band. The remaining channels will be assigned to the 27- to 51-mc band for transmission from the studio to the school tap-off. At that point they will be converted to standard channels 7, 9, and 11.

Transmission Considerations

A successful TV distribution system must be capable of accepting video and audio signals at its input and delivering these signals to all the viewing locations with a minimum degradation of the information being transmitted. Furthermore, it must be capable of performing this function on a continuous basis since even short service interruptions can seriously upset a tightly planned curriculum In practice the real test is whether or not the transmission medium is capable of providing a video and audio presentation which will meet the requirements set forth for efficient teaching. To meet these requirements the Bell System has set up the following specifications for a system of this type.

1. *Radiation.* Prior to the final choice of channel assignments in any closed-circuit TV system consideration must be given to any nearby TV station in operation. In general it may be assumed that the distribution system itself can be shielded sufficiently in order to eliminate any objectionable outside signal. However, at the viewing end of the system, the problem is somewhat more difficult. A typical TV set in the classroom, even without benefit of an antenna, is likely to experience interference from any nearby station operating on the same frequency as used for the closed-circuit program. In addition, interference from certain stations operating on cochannels can also become a problem. In such cases actual tests in the classrooms are essential.

Also to be considered is radiation from the closed-circuit equipments and cables. The limits in this case are prescribed by the Federal Communications Commission. Difficulty should not be encountered with normal transmission levels and terminations.

2. *Signal Levels.* A typical amplifier output for the video carrier would be +40 dbmv where zero dbmv represents 1 millivolt across 75 ohms. Incidentally, the impedance of all the equipments and cables used in this system is a nominal 75 ohms. The levels of the audio carriers are adjusted to be 6 db lower than the video carriers. This arrangement reduces the load-carrying requirements for the amplifiers without noticeably increasing the noise level in the audio. Should carrier levels be permitted to exceed about +44 dbmv excessive overloading and intermodulation may result.

3. *Signal-to-Noise Ratios.* Tests indicate that if the minimum carrier level in the transmission system is not permitted to drop below +8 dbmv, noise will not become objectionable. Actually, once signals leave the transmission system and enter the school distribution system, levels are permitted to fall to about zero dbmv at the TV sets. At this point the carrier-to-noise ratio should not fall below 46 db. It follows that this ratio must be greater than 46 db at all points between the studio and the most distant viewing location.

Fig. 5. Line amplifier with cover removed

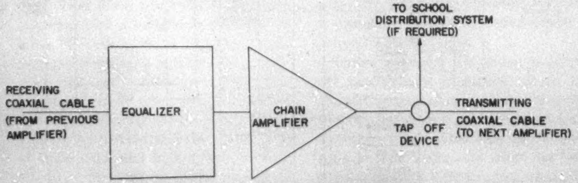


Fig. 4. Block diagram of line amplifier

4. *Cross Modulation.* This is primarily a function of: 1. the maximum number of amplifiers in tandem, 2. the maximum transmitted level, and 3. the number of channels in operation at a given time. The maximum number of amplifiers encountered between the studio and the most distant classroom in this installation is 30. Assuming normal transmission levels and a common synchronous source for all programs originating at the studio, interference should not be objectionable. With up to three channels in operation, cross modulation on the Cortland County system has not been a problem.

5. *Phase.* Since distribution is at radio frequencies, phase is not a problem in this system.

Transmission Details

At the studio location the telephone company receives video and audio signals at levels of 1 volt peak to peak and zero volume unit, which are fed via the patch panels to the transmitting terminals. These convert the signals to their assigned channels. As mentioned earlier, this terminal is actually a miniature TV transmitter which receives audio and video signals and converts them to a standard composite signal at a predetermined radio frequency in the 27- to 88-mc band. The several composite TV signals and the 77-mc pilot frequency, which is generated in the telephone company's equipment bay, are combined into a single output. This is then fed to an amplifier and transmitted out on the coaxial cable. The

Vol.
Pt. 1
1959

channelization and transmission equipment bay is shown in Fig. 3.

Amplifiers are spaced along the coaxial cable at intervals which will maintain the signal levels within the limits indicated earlier. The locations along the coaxial cable are dependent upon the following:

1. Loss in the coaxial cable at the highest operating frequency.
2. Minimum acceptable level at the input of amplifiers.
3. Additional losses due to any networks, filters, etc., inserted in amplifier sections.

Two types of amplifiers are used in this installation. One, called a line amplifier, simply provides gain which is manually adjustable from about 16 to 28 db. Figs. 4 and 5 illustrate the installations. About two thirds of the amplifiers along the cable are of this type. The remaining amplifiers are known as automatic level control stations (ALC) and are discussed in detail. Both the line amplifier and the ALC station include a chain or distributed type of amplifier unit, which is especially suited for this service since its operation is practically independent of tube failure. This feature is a basic characteristic of the chain amplifier which utilizes 12 tubes each contributing an equal amount of gain throughout the band. Loss of an individual tube has practically no effect on picture quality. With the type of cable used in this system the section preceding a line amplifier can be extended to about 4,000 feet without causing levels to drop below +8 dbmv. It is necessary to maintain a higher signal level at the input of an ALC station. Therefore the preceding section is limited to about 2,700 feet. Both the highest and lowest anticipated air temperatures must be considered in determining amplifier spacings. In summer, when cable losses are at a maximum, the levels must not be permitted to drop so low as to cause excessive noise. Conversely, in winter, when cable losses are at a minimum, the levels must not be permitted to reach a point where intermodulation and overloading become excessive. In addition to the level changes resulting from these temperature variations, a "tilting" of the coaxial loss-frequency characteristics also occurs. Therefore seasonal equalization variations become an important factor in the control of signal levels.

To assure acceptable picture quality, it is essential that the signal levels be maintained within the extremes just described. This is accomplished as follows:

1. ALC stations are provided so as to compensate for day-to-day temperature variations.
2. Seasonal gain and equalization adjustments are necessary to compensate for major temperature excursions.

At about every third amplifier location, or 81 db of cable loss at the highest operating frequency, an automatic level control station is provided; see Figs. 6 and 7 for block diagram and typical installation. The purpose of the ALC station is to compensate for cable loss variations brought about by daily changes in temperature. Following the chain amplifier unit at the ALC station is a filter which has a single input and two outputs; one output passes the entire 27- to 88-mc band on to the transmitting coaxial cable, while the other has a narrow band-pass characteristic centered around 77 mc. The latter diverts some of the 77-mc pilot signal to a specially provided amplifier which convert this signal to a d-c bias voltage of a value proportional to that of the pilot signal level. This bias voltage is used to control the gain of the chain amplifier at an ALC station. As the temperature of the cable varies, the cable loss and the pilot level fluctuate accordingly. Since the value of the bias voltage is a function of the pilot level, it is possible to maintain the amplifier output at a fairly constant value even though temperatures and cable loss may vary over a fairly wide range. A single ALC is capable of handling the level variation normally encountered in the cable section back to the preceding ALC station. With cable of the type used in this installation the distance between ALC stations or from the origination point to the first station can be extended to about 10,700 feet. The level adjustments provided by the ALC stations are flat gain in nature. Compensation for seasonal variations of the frequency and flat gain characteristics of the cable must be accomplished by manual gain adjustments and equalizer changes. This becomes necessary when the change in temperature is great enough to cause excessive loss and tilt.

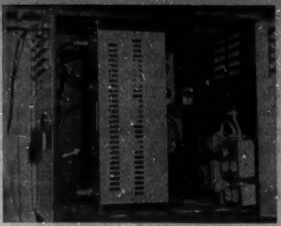


Fig. 7. Automatic level control amplifier with regulator unit mounted at rear

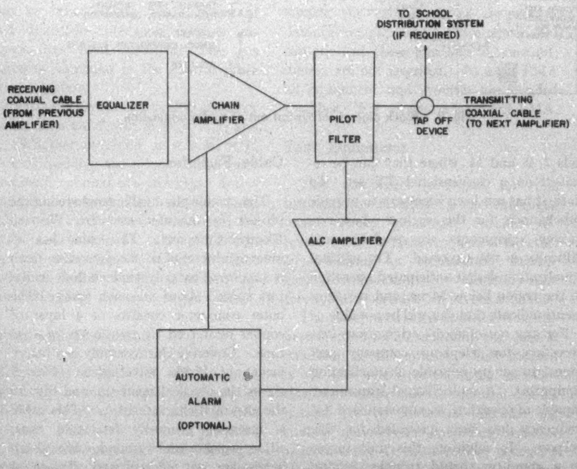

TO SCHOOL DISTRIBUTION SYSTEM (IF REQUIRED)
RECEIVING COAXIAL CABLE (FROM PREVIOUS AMPLIFIER)
EQUALIZER
CHAIN AMPLIFIER
PILOT FILTER
TAP OFF DEVICE
TRANSMITTING COAXIAL CABLE (TO NEXT AMPLIFIER)
ALC AMPLIFIER
AUTOMATIC ALARM (OPTIONAL)

Fig. 6. Block diagram of automatic level control amplifier

At school locations, a tap-off is required for feeding the distribution system. This device must not only pass all frequencies in the band used but must also provide isolation between the main line and the school room distribution system. The latter is essential since without isolation any sort of trouble in the distribution facilities not only could seriously degrade the signal for other locations but under certain conditions could cause complete failure of the system. The tap-off device used for this installation inserts a loss of less than 1 db in the main line, 10 db in the school feeder, and provides 26 db of isolation. This tap-off unit uses the directional coupler principle and requires no special equalization for the band in use.

Out-of-Studio Originations

It will be noted in Fig. 1 that an amplifier is located in most school buildings. While this does not permit utilization of maximum repeater spacings, it does provide considerable flexibility. Since this system is experimental in nature the arrangement appears well justified. This flexibility is especially desirable when out-of-studio originations are considered. With the round-robin it is possible for any school to be set up as a studio. For a typical illustration of this feature assume channel 6 and the pilot originating in the studio in the normal manner, and channel 2 originating in a special studio set up in the high school; see Fig. 8. Under normal operating conditions where all programs originate at the studio, the return leg of the cable is terminated at pole-mounted amplifier (PMA) 8. For this special case, however, it is necessary to close the loop at PMA 8 and permit passage of signals originated at the high school, but to block any signals originated in the regular studio. These latter signals upon reaching PMA 8 have already been viewed at each school location. Should they be permitted to re-enter the system a "singing" condition would exist and normal transmission would not be possible. At the high school, channel 2 would be blocked in order to prevent re-entrance and the singing condition. To prevent a build-up of circulating a noise on the closed loop, filters for blocking unused portions of the band (i.e., 27- to 54-mc, plus channels 3 and 4 and the region above channel 6) are also provided.

The blocking and passage of the various signals is accomplished through insertion of band elimination filters at the appropriate locations. Filters having characteristics sufficiently sharp to permit adjacent channel operation and still provide the necessary pass or block bands are not available. Therefore assignment must be confined to alternate channels 2, 4, and 6. Should more than three channels be required the 15- to 54-mc band will be used. Since the standard TV set does not receive these frequencies it is necessary that the telephone company provide converters at each school to translate the three subchannels to channels 7, 9, and 11, where they can be received on a conventional TV set. To date, it has not been necessary to provide subchannels for this system; however, proven equipments are available and difficulty is not expected. The original equalization design anticipated operation in the region below 54 mc and measurements indicate that this will be possible.

For any out-of-studio originations it is necessary for telephone company personnel to set up portable channelization equipment. A multichannel transmitter capable of operating on any standard TV frequency has been provided for this purpose. In addition, this unit serves as a stand-by should trouble develop in any of the regular transmitters.

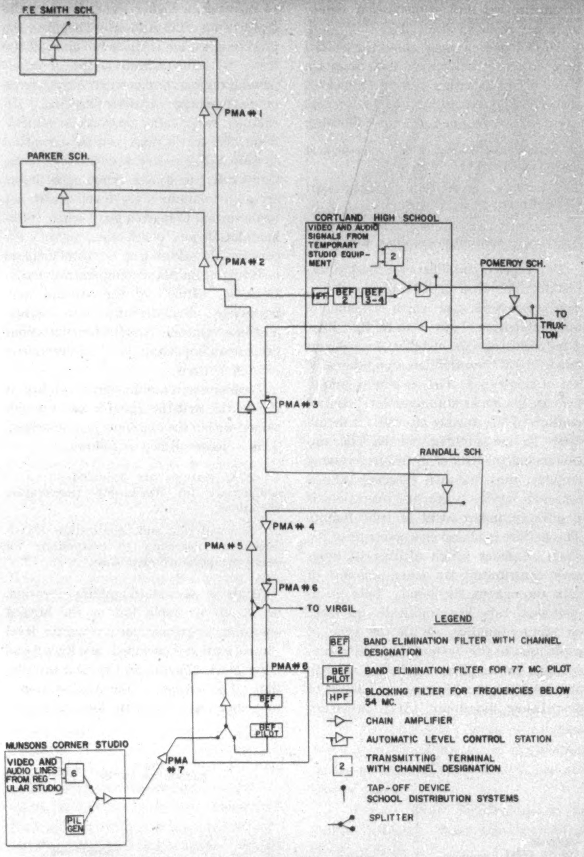

F.E SMITH SCH.
PMA #1
PARKER SCH.
PMA #2
CORTLAND HIGH SCHOOL
VIDEO AND AUDIO SIGNALS FROM TEMPORARY STUDIO EQUIPMENT
2
HPF
BEF 2
BEF 3-4
POMEROY SCH.
TO TRUXTON
PMA #3
RANDALL SCH.
PMA #4
PMA #5
PMA #6
TO VIRGIL
PMA#8
BEF
BEF PILOT
MUNSONS CORNER STUDIO
VIDEO AND AUDIO LINES FROM REGULAR STUDIO
6
PIL GEN
PMA #7
LEGEND
BEF 2
BAND ELIMINATION FILTER WITH CHANNEL DESIGNATION
BEF PILOT
BAND ELIMINATION FILTER FOR 77 MC. PILOT
HPF
BLOCKING FILTER FOR FREQUENCIES BELOW 54 MC.
CHAIN AMPLIFIER
AUTOMATIC LEVEL CONTROL STATION
2
TRANSMITTING TERMINAL WITH CHANNEL DESIGNATION
TAP-OFF DEVICE SCHOOL DISTRIBUTION SYSTEMS
SPLITTER

Fig. 8. Block diagram of typical out-of-studio origination

Cable Facilities

The outside plant cable employed in the project is manufactured by Western Electric Company. This cable has an inner conductor of no. 9 copper wire which is supported by polyethylene disk insulators spaced about one inch apart. The outer conductor consists of a layer of copper reinforced by two layers of steel tape. Covering this assembly is a heavy jacket of black polyethylene. Fig. 9 shows the physical make-up, and Fig. 10 the electrical characteristics. This cable is intended primarily for main runs, either aerial or underground, where sharp bends are not encountered. In aerial installations it may be lashed to a specially

Vol.
Pt. 1
195

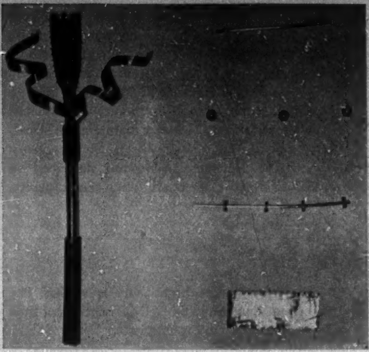


Fig. 9 (left). Exploded view of Western Electric Company's 3/8-inch air dielectric coaxial cable

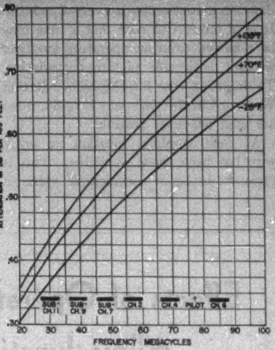


Fig. 10 (right). Attenuation characteristic for 3/8-inch air dielectric coaxial cable

provided messenger or along with an existing cable. Splicing is accomplished through the use of sleeves on both inner and outer conductors. At amplifier locations or at any point where considerable bending is required the cable is terminated in a coaxial plug which is equipped with a male N-type connector. From this point a more flexible solid dielectric cable is used. The cable used for these extensions is similar in electrical characteristics to *RG 11U* but it includes a double shield. Because of its solid dielectric the more flexible cable has greater loss than the main cable. This is not serious, however, since its use is confined to short sections.

For the main cable runs gas pressure has been provided. At amplifier locations or other points where pieces of solid dielectric cable are used a gas bypass has been installed. The gas source in Cortland is the central office dry-air system.

Equipment layouts for typical amplifier installations are shown in Figs. 5, 7, and 11. Whether mounted on a pole or in a school building the same weatherproof equipment cabinets are provided. Interconnection between the various units within the cabinet is accomplished by patch cords made of cable similar in electrical characteristics to *RG 59U* but having a double shield. Standard N-type connectors are used in all cases.

Power and Protection

From Figs. 5 and 7 it will be noted that both regulated and nonregulated 115-volt outlets are installed in the cabinets. Regulated power is provided for the system amplifiers while nonregulated power is provided for test equipment, soldering irons, etc. In a pole-mounted location, an unmetered power line covered by a fiber shield is extended from the 115-volt line overhead down to the telephone company space on the pole. This portion of the work is handled by the power company. From this point an electrician has installed conduit and extended the power conductors directly through the bottom of the amplifier cabinet. Power as well as all the coaxial cables enter the bottom of the equipment cabinets to avoid drilling in the tops which might result in water leakage. Once inside the equipment cabinet, the power leads are wired directly to a circuit breaker. For lightning and surge protection a pellet protector is bridged across the incoming power line. This protector is mounted outside the cabinet so as to prevent damage to equipment. Since power consumption at these locations is constant, meters are not required. At each PMA all equipment and grounds are bonded together. Fig. 11 shows a typical PMA.

Test Equipment

Test equipment provided for this installation included a field-strength meter, sweep generator, variable attenuator, and a radio-frequency detector. In addition, an oscilloscope, marker generator, vacuum-tube voltmeter, tube tester, and a portable TV set are available.

The sweep technique appears to be the best method for determining the general condition of the system. This method is not only useful for the initial installation work where various equalizer combinations are to be tested but also for routine maintenance including seasonal adjustments. It is, of course, necessary that the system be out of service during periods when this method is employed. For in-service testing, the levels of the TV carriers may be measured at the monitor jacks of any amplifier.

Spare Equipment

To assure reasonable service continuity, at least one spare of each type of operating equipment is provided. Provision of spares of the more common items such as line amplifiers, vacuum tubes, connectors, and equalizers is considerably more liberal than that of lesser used items. This stock of spare equipment along with test equipment is stored in a central location under direct control of the service foreman responsible for system maintenance.

Maintenance

Maintenance to date on this system has been primarily a matter of testing tubes in the chain amplifiers and it is

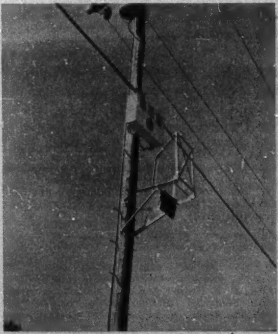


Fig. 11. Typical PMA installation

done on a routine basis every 4 months. It is of interest to note that a tube may be removed from an operating amplifier without noticeable picture degradation. To improve over-all reliability, further Western Electric Company long-life tubes are used wherever possible.

At each amplifier location an order wire has been made available. This is expected to be of considerable value in maintenance procedures. Also an alarm unit is provided at strategic locations to monitor the level of the pilot signal, which will operate should any departure exceed certain specified limits.

Conclusions

Detailed tests have been made at the studio and at the various school locations. These indicate that the pictures delivered to the classroom TV sets were essentially equivalent to those on the studio monitors. The system has been in operation since September 1958, and to date no change in picture quality has been noted. From a maintenance standpoint there have been no problems. The transmission system described in this paper meets all design requirements.

Effects of Operation of Germanium Alloy Junction Transistors Above Rated Conditions

B. C. SPRADLIN
ASSOCIATE MEMBER AIEE

As a part of the investigations and analyses performed at Battelle Memorial Institute for Remington Rand Univac, some exploratory experiments were performed at various operating conditions with type *ZJ1-501*(*4JD1A-62*) and *ZJ11-501*(*4JD1A-70*) transistors. Both of these transistors are special type *2N123* p-n-p alloy junction high-frequency switching transistors. (The primary physical difference between the *ZJ1-501* and *ZJ11-501* transistors is the thicker base region of the *ZJ1-501*.) These exploratory experiments were performed as a portion of a program directed toward the development of accelerated life test procedures for transistors.

Experimental Procedures

The two types of transistors were subjected to the operating conditions listed in Table I. Since both types of transistors are rated at 150 milliwatts power dissipation in an ambient temperature of 25 C (degrees centigrade), all the operating conditions of the experiment are above the transistors' ratings.

The transistors were operated in common-emitter circuits as shown in Fig. 1 for approximately 1,000 hours.

The base resistor was a 1% carbon-film resistor of 10,000 ohms, and the base voltage was 3 volts for all the transistors. A 5% wire-wound resistor R was employed at less than one-fifth its rated power dissipation in the collector-emitter circuit. The resistance value for this resistor and the collector-emitter voltage supply E_{ce} varied according to stress condition and transistor type.

Transistors, as is the case for any component, vary in characteristics from one unit to another. For this reason, the power-dissipation levels as shown in Table I do not signify that all of the transistors under a certain heading dissipated exactly that power. Further, the power dissipations were not held constant with time. Instead, the supply voltage in the collector-emitter circuit was set initially at the ambient temperature for the stress condition, so that the median power dissipation for the group of 20 transistors was equal to the power dissipation that was desired for each stress condition. This voltage, then, was held constant throughout the experiment. Since the collector resistor was a wire-wound type and was loaded at only a fraction of its rating, it did not change appreciably with time. Measurements of the collector-emitter power dissipation were made three times a week. The median values plotted as functions of time are shown in Figs. 2 and 3.

Room temperature measurements of the transistor parameters were made initially six times during the experiments, and after approximately 1,000 hours of operation. Checks for open or short circuits were made daily, except for week ends. Graphs of the median values of transistor parameters as functions of time are presented in Figs. 4–6. Discontinuities in these curves are caused by equipment or instrumentation malfunctioning. It appears that these effects can be neglected in considering the over-all effects of operation at the various stress conditions. Points at which the curves become vertical lines indicate that one-half or more of the transistors operating at that particular stress condition have failed.

Measurements for small-signal gain h_{fe}, d-c gain B, punch-through voltage V_{pt}, and breakdown voltage BV_{CBO} were made on an American Electronics Laboratories,

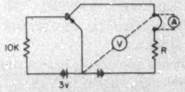


Fig. 1. Circuit employed for transistor exploratory experiments

Table I. Factorial Representation of the Operating Conditions of the Exploratory Experiments

Ambient Temperature, C	Power Dissipation, Milliwatts		
	100	150	200
Type ZJ11-501(4JD1A-62) Transistors			
100	20	20	20
120	20	20	20
Type ZJ11-501(4JD1A-70) Transistors			
80	20	20	20
90	20	20	20
100	20	20	20

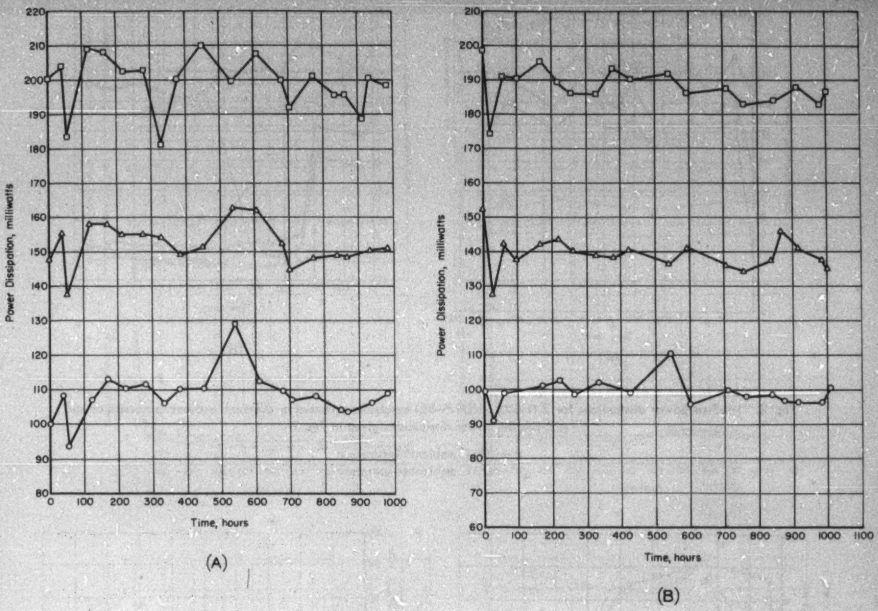
Note: The number 20 is the number of transistors operated at each combination of temperature and power dissipation.

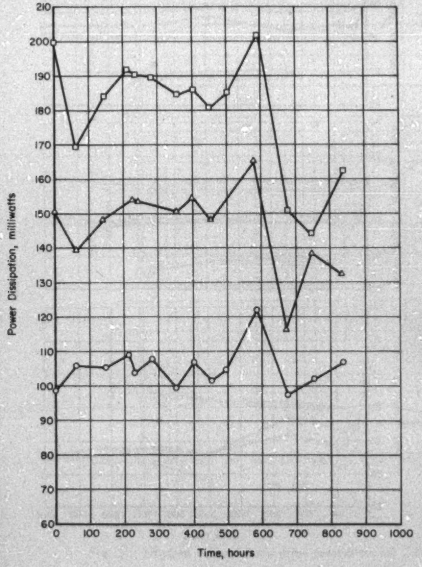
Paper 59-145, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 3, 1958; made available for printing December 1, 1958.

B. C. Spradlin is with the Battelle Memorial Institute, Columbus, Ohio.

The author wishes to express his appreciation for the advice and co-operation received from H. M. Braner, K. E. Hassler, and other staff members of Batelle Memorial Institute and personnel from Remington Rand Univac. Part of this work was done for Remington Rand Univac under subcontract *68944* of Air Force contract *AF 04(6450-20)*.

Vol.
Pt. 1
1959





(C)

Fig. 2. Median power dissipations for ZJ11-501 (4JD1A-70) transistors operated in different ambient temperatures

A—80 C ambient temperature
B—90 C ambient temperature
C—100 C ambient temperature

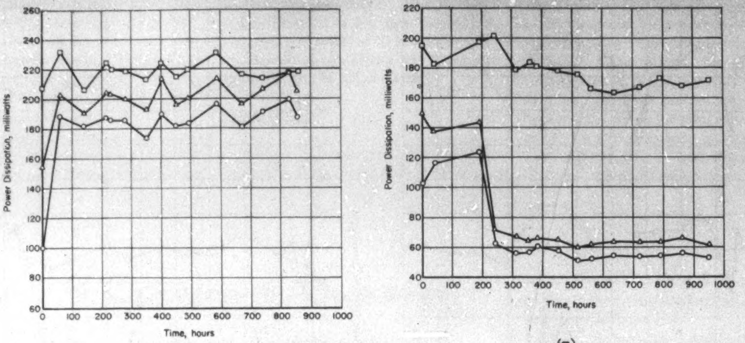
Power-dissipation levels:
○ 100 milliwatts
△ 150 milliwatts
□ 200 milliwatts

Inc., model *126H* transistor curve tracer with a Dumont model *340* oscilloscope. The small-signal gain and d-c gain were calculated from displays of the R_{22} family of curves (collector current versus collector voltage for various base currents). The collector currents were read for a collector voltage of 1 volt and 0, 2, 4, and 6 microamperes base current. The curves that are given are curves of the median values of d-c gain and small-signal gain with 6 microamperes in the base and a change of 4 to 6 microamperes base current, respectively. A typical R_{22} family of curves is shown in Fig. 7.

A typical display of curves of punch-through and breakdown voltages is shown in Fig. 8. The breakdown voltage curve was aligned with the horizontal "zero" line and a reading was made of where the curve leaves this line. Punch-through voltage was read where the punch-through curve crosses the horizontal 1-volt line, as illustrated.

Measurements of I_{CO} and I_{EO} were made with a Radio Corporation of America model *WV-84A* microammeter and a d-c transistor power supply. Configurations for these measurements are shown in Fig. 9.

(*Continued on page 352.*)



(A)

(B)

Fig. 3. Median power dissipations for ZJ1-501 (4JD1A-62) transistors operated in different ambient temperatures and at the initial power-dissipations given in Fig. 2

A—100 C ambient temperature
B—120 C ambient temperature

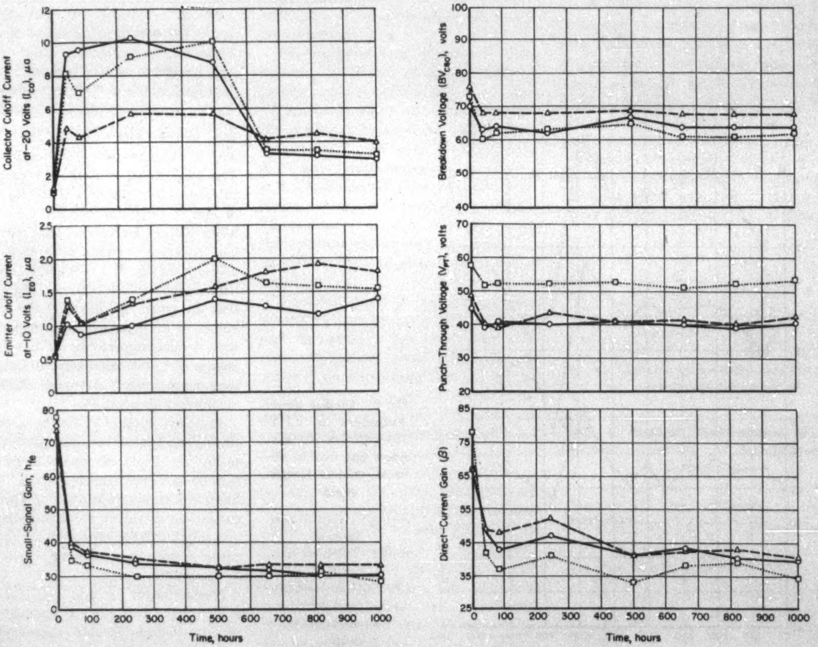


Fig. 4. Median values of transistor parameters for ZJ11-501 (4JD1A-70) transistors operated in different ambient temperature of 80 C and at the initial power dissipations given in Fig. 2 caption

Vol.
Pt. 1
1959

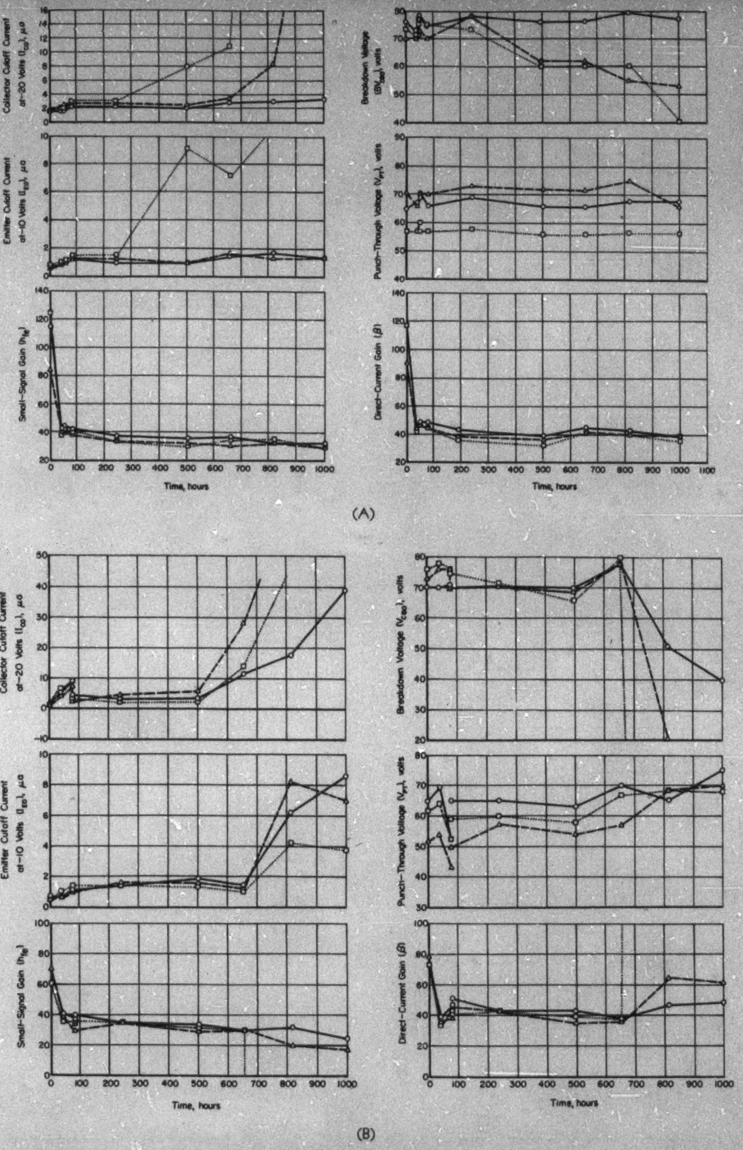


Fig. 5. Median values of transistor parameters for ZJ11-501 (4JD1A-70) transistors operated in different ambient temperatures and at the initial power dissipations given in Fig. 2 caption

A—90 C ambient temperature
B—100 C ambient temperature

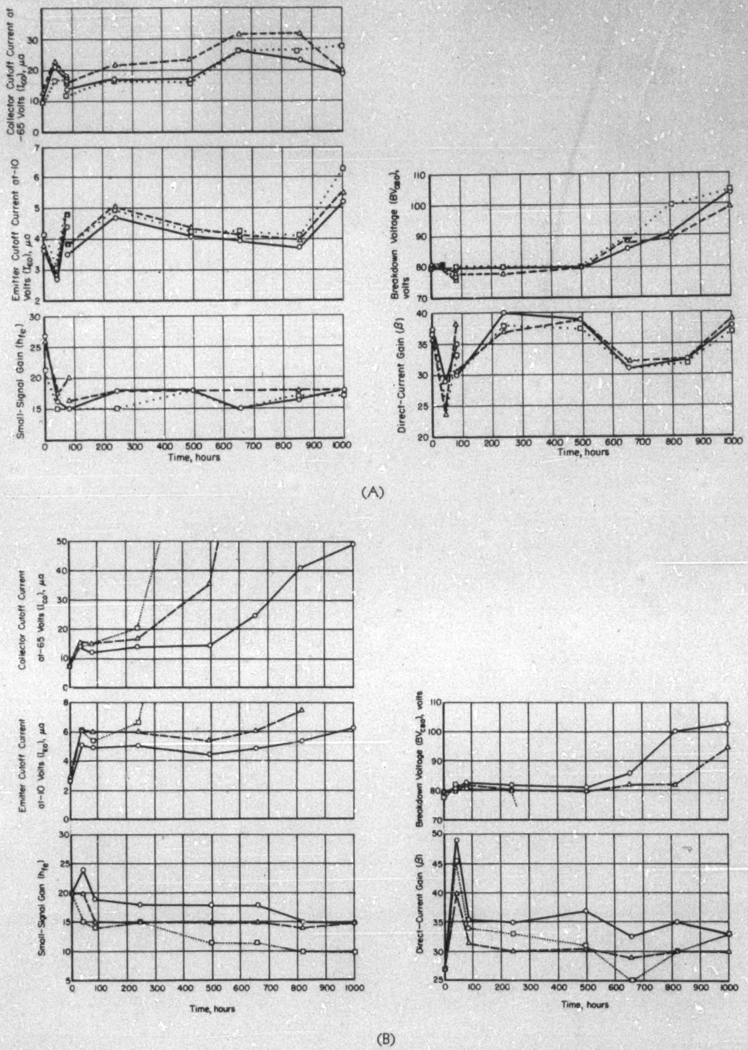


Fig. 6. Median values of transistor parameters for ZJ1-501 (4JD1A-62) transistors operated in different ambient temperatures and at the initial power dissipations given in Fig. 2 caption

A—100 C ambient temperature
B—120 C ambient temperature

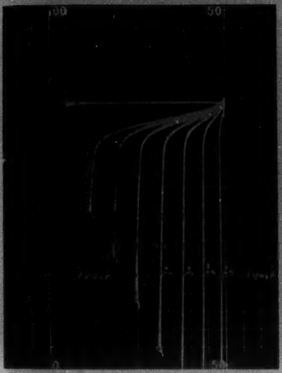


Fig. 7. Typical R_{cc} family of curves for transistors employed in exploratory experiments

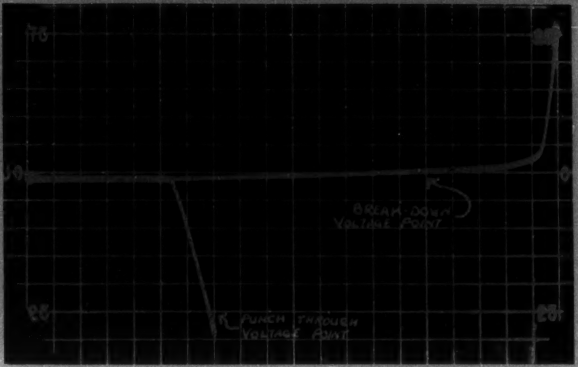


Fig. 8. Punch-through and breakdown voltage display

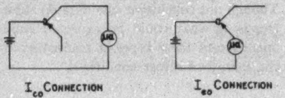
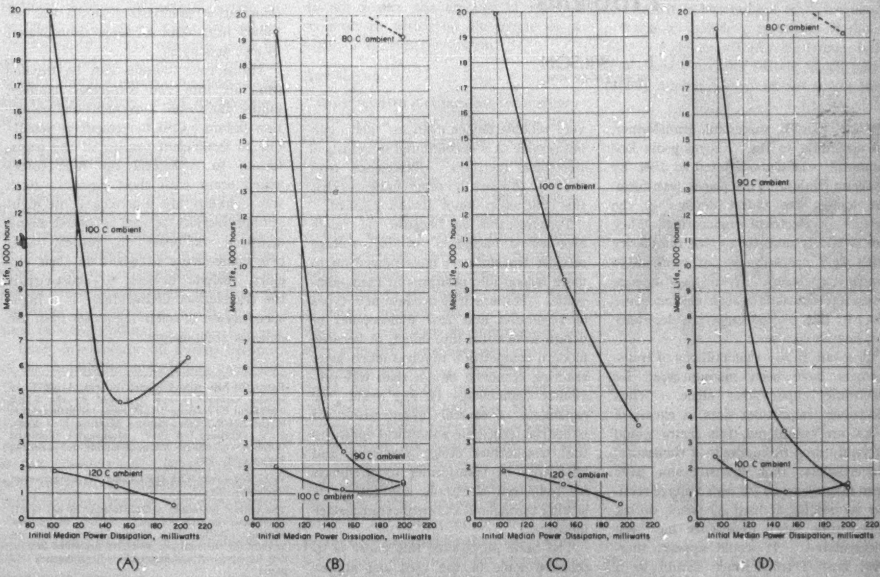


Fig. 9 (left). Circuit configurations for I_{CO} and I_{EO} measurements

Fig. 10. Mean lives of type ZJ1-501 transistors as a function of median power dissipation for different ambient temperatures

A,B—Before arraying
C,D—After arraying



(A) (B) (C) (D)

Effects of Various Stress Conditions by Comparison of Mean Lives

The mean lives of the transistors for each stress condition were calculated to compare the effects of operation at the various stress conditions. These mean lives were computed from the formula

$$M=\frac{\text{aggregate operating hours}}{\text{number of failures}}$$

The mean life is the best estimate from the data available of expected life of a randomly selected transistor operated at the specified stress condition.

The failure criteria used for each type of transistor were:

ZJ1-501 (*4JD1A-62*) transistors, inoperative or I_{CO} at −65 volts greater than 200 microamperes

ZJ11-501 (*4JD1A-70*) transistors, inoperative or I_{CO} at −20 volts greater than 100 microamperes

The mean lives as a function of median power dissipation for different ambient temperatures are presented in Figs. 10(A) and 10(B). It can be seen that there appears to be an anomaly in the 100 C curves from 150 milliwatts to 200 milliwatts for both types of transistors, for these curves indicate that 150 milliwatts power dissipation is a greater stress than 200 milliwatts. It must be remembered, however, that the power dissipation shown is the median initial power dissipation for a group of 20 transistors. Because of this, the ranges around the median power dissipation overlapped between two adjacent dissipations of the factorial. To correct for this overlap, the initial power dissipations for the 60 transistors at each ambient temperature were arrayed from the least to the largest initial dissipation and then divided into three goups. The mean lives of the transistors for each of these three groups were computed and plotted as functions of the median initial power dissipation for each group. These graphs are shown in Figs. 10(C) and 10(D). It can be seen that the median initial power dissipation after arraying is approximately the same as it was before arraying. Further, the 100 C curve for the *ZJ1-501(4JD1A-62)* transistors is now more like what would be expected. The 100 C curve for the *ZJ11-501(4JD1A-70)* transistors still appears to have an anomalous portion. However, if confidence limits are placed on the mean lives and the small sample size is considered, the curves are not unreasonable.

Conclusions

The exploratory experiments used for two similar types of alloy junction transistors have been described and the median parameter variations with time have been presented in graphical form. Further, estimates of the mean lives of the transistors operating at the stress conditions of the exploratory experiments have been given. The data presented in the foregoing are not conclusive and are presented merely to show indications of the effects of operation at the various stress conditions employed in the exploratory experiments. Though not conclusive, the data do show trends of what could be expected from operation of these types of transistors at the described stress conditions.

Vol.
Pt. 1
1959

High-Voltage Rectifier Transformer Problems

T. L. WILSON
NONMEMBER AIEE

THE FIRST successful transformer, according to the "Encyclopedia Britannica," was demonstrated in 1886 by William Staley. This places transformers among the oldest devices in the science of electrical engineering. They are so widely used today that familiarity with their operation is not confined to trained engineers. Even the average housewife is aware of their existence and the part they play in supplying electricity for household uses.

It is safe to say that millions of transformers have been manufactured for electronic applications alone. When electronic transformer sizes in excess of 1 kw are considered, this figure would probably drop to hundreds of thousands. Nevertheless, such transformers have been manufactured for high-voltage rectifier service for at least 40 years, during which time much experience has been accumulated. It would appear, therfore, that a transformer should be a very reliable device even for such rigorous service as is encountered in industrial electronic rectifiers. Experience over the past 15 years, however, indicates that this conclusion is not always justified.

Perhaps this is because the more experienced designers of high-voltage rectifier transformers have moved on to more responsible positions in their companies. Perhaps it is because new types of insulation and new philosophies of design have come into effect, or because modern competition requires much more exacting economic design than was previously demanded. In any event, the experience of several manufacturers of industrial electronic equipment indicates that transformer design technique and knowledge of transformer requirements for industrial electronic applications is lacking throughout the entire transformer industry.

The basic purpose of this paper is to describe some of the uses and abuses that high-voltage industrial transformers encounter in rectifier service and to partially analyze the reasons for some of the failures that result. In most cases, the information summarized herein was obtained from data on large transformers (5-kva and over).

Transformer failures can be divided generally into two categories, namely, failure due to coil movement and insulation failure. Coil movement is usually due to insufficient bracing of the transformer to withstand the high forces which occur with short circuits. Such short circuits are unavoidable in high-voltage rectifier service. Indeed, many designers, in order to protect delicate components elsewhere in their circuits, include devices to short-circuit the transformer deliberately, as a high-speed means of removing power from the delicate components.

Paper 59-150, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 15, 1958; made available for printing January 30, 1959.

T. L. Wilson is with the Chemetron Corporation, Louisville, Ky.

The author gratefully acknowledges the aid of his colleagues in the Girdler Process Equipment Division of Chemetron Corporation for their aid in thinking through the problems described herein, and for their critical suggestions in preparing this paper.

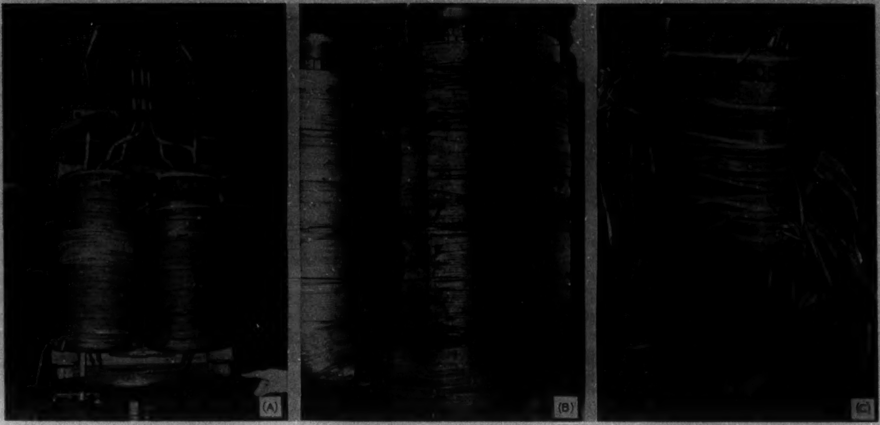


Fig. 1. Coil movement failure

A—Core and coil assembly of liquid-cooled transformer after typical failure
B—Close-up view of another failure
C—One of the coils from (A) cut to show insulation failure

1. 78
1
59

Insulation failure is due to insufficient insulation to withstand the high-voltage transients which occur in the rectifier service. These transients are sometimes troublesome in thyratron rectifiers, although they also occur in ordinary rectifiers because of several types of secondary switching in which the load on the transformer may suddenly drop to zero. In rectifier service it is possible to get a simultaneous combination of the two types of failures, as will be described later.

Coil Movement Failures

Experience has shown that failure due to coil movement has been most prevalent in liquid-cooled transformers. Because of safety considerations, the liquid is often of the noninflammable type. The use of such liquids limits the type of varnishes and insulation which may be used in a transformer.

Even in ordinary oil types, because of the desirability of including ducts for the flow of oil through the windings, impregnation cannot bind the high-voltage and low-voltage windings together as effectively as it can in air-cooled transformers. In one case, of nine 100-kva transformers, all nine had coil movement failures and were rebuilt within a 1-year service period. Obviously, such experience is expensive to the transformer manufacturer, but is even more expensive to the equipment manufacturer because of the cost of service calls and the possible loss of business due to loss of reputation for his product, and to the user because of lost production. It is surprising that transformer manufacturers apparently do not realize the amount of bracing required to withstand the rigors of a high-voltage rectifier system.

Arc Backs

Most modern rectifiers use the mercury vapor type of tube. These tubes have quite a limited operating temperature range and a rather low maximum voltage. Equipment manufacturers often include components for controlling the temperature of the condensed mercury in their rectifier tubes. They also use the tubes at considerably lower than maximum rated voltage. Even then, rectifiers are subject to arc backs, which in some industrial services may occur once a week or sometimes as infrequently as once a month; in other cases arc backs may occur more often than once a day. Some are of short duration and are self-clearing; these will give the transformer a physical force shock which can be heard as an audible thud. Other arc backs (probably more severe) are well established in the tube and exist until the power line protection system removes the source of power. These produce an even louder thud followed by a loud hum attesting to the intensity of the blow the windings have received. As will be explained later, arc backs occur very rapidly and can cause other types of failures as well as coil movement.

Crowbar

In both communication and industrial services, it is becoming a more frequent practice to include a so-called crowbar. This is a device, usually electronic, but sometimes a vacuum switch, which very rapidly applies a short circuit across the

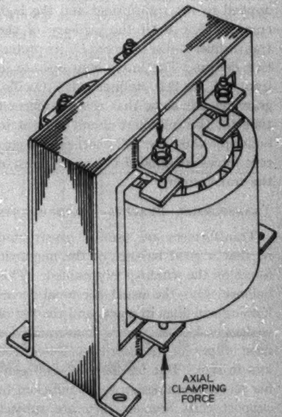


Fig. 2. Sketch of C-clamp bracing of two windings of typical transformer

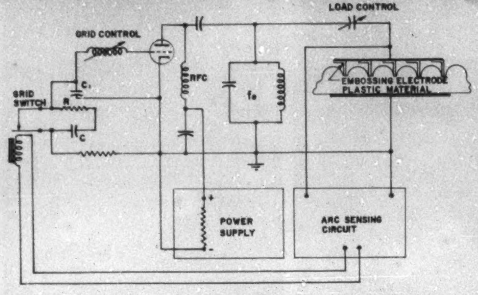


Fig. 3. Experimental oscillator circuit with arc-stopping device connected in grid circuit

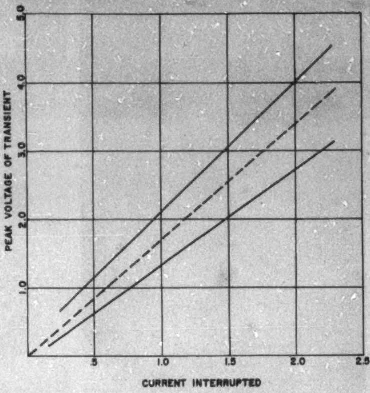


Fig. 4 (right). Curve showing maximum and minimum value of rectifier output transient versus current interrupted by means of grid switch of Fig. 3

d-c output of a rectifier system. The short circuit is applied in times varying from a few microseconds to a few milliseconds when a fault occurs in the r-f (radio-frequency) system. The purpose is to remove power from the r-f system, particularly the vacuum tubes, in as short a time as possible. In communications service crowbars must operate very rapidly so that the large quantity of energy stored in filter condensers will not be discharged into the fault which may occur in the delicate grid or seal structure of a vacuum tube. In industrial equipment, filter condensers are usually not used. Therefore, slightly slower-acting devices are satisfactory.

In either case, the short circuit is applied to the transformer and the high currents flow until the primary of the transformer is disconnected by its protective device. The latter may operate in from 17 to 200 milliseconds. Oscillogram records show that this is sufficient time for the full short-circuit current to buildup; thus the full short-circuit current forces are applied to the windings of the transformer.

TRANSFORMER BRACING REQUIREMENTS

Transformers are usually constructed so that a good balance of the magnetic forces on the windings is obtained. This balance, plus the usual frictional forces holding windings in place, and the use of wedges and other bracing, is adequate for most types of transformers below 100 kva in size. But for high-voltage rectifier service such design is not sufficient to support coils because they are subject to frequent short circuits.

Figs. 1(A) and (B) show the results of the use of normal distribution transformer construction in rectifier service. The crushing of the margin insulation and the bending of the coils are the result of very intense forces acting to move the outer windings downward. Fig. 1(C) shows the electrical failure at the point of crushing of the margin insulation.

Very strong bracing is required. Experience shows that the only successful design is one in which several screw-type clamps are placed around the periphery of the winding, squeezing the coils together in the axial direction. This "C-clamp" bracing is so designated because the windings are effectively placed in rigid, strong C-clamps. Fig. 2 shows the C-clamp type of bracing. Relative movement of the high-voltage and the low-voltage winding is prevented by the rigid clamping action.

It is well known that magnetic forces which tend to move the windings axially, can be balanced quite well (although not completely) in the original design and construction of the transformer. However, if the windings have a chance to move even slightly, this balance is rapidly destroyed. This increase in force, of course, can cause the poorly braced winding to move more on the next short-circuit blow and the additional motion causes further net force until a runaway condition is obtained. The resulting force is terrific, as from a sledge hammer, and completely crushes the margin insulation of the transformers. C-clamp or other adequate bracing, if properly applied to a well-balanced design, will prevent that first movement of the coils and thus prevent the destruction of the original force balance.

Table I. Results of Varying R-C Network Capacity

Capacitance, Microfarads	Maximum E, Centimeters	Minimum E, Centimeters	Maximum Time, Milliseconds	Minimum Time, Milliseconds
12	4.0	3.2	50	44
4	3.8	3.0	26	22
1.0	4.2	3.4	15	12.4
0.25	4.2	2.9	13.6	11.0
0.1	4.3	3.2	12.2	10.0

To emphasize, adequate bracing must be included in high-voltage rectifier transformers. It does not add appreciably to the cost of manufacture of the transformer, inasmuch as the clamps, for example, are simple devices, usually welded to the frame of the transformer. When one considers the cost of a transformer failure, not only to the supplier but to the equipment manufacturer and his customer, the investment in adequate bracing is very well justified. In the author's opinion, regardless of how small the kva of a transformer, particularly for liquid-cooled rectifier transformers, the C-clamp type of bracing should be specified and used in every case.

Insulation Failure

There are several characteristics of rectifiers which can lead to insulation failure in the transformers used with them. Some of these are perhaps more frequently encountered in industrial equipment than in communication equipment, but they all arise from the well-known characteristic of inductances in which a sudden interruption of current will cause a high voltage to appear across the terminals. This voltage can easily be destructively high and will not necessarily be evenly distributed throughout a given inductance such as a transformer winding.

Arc Interrupters

There are several ways in which the current flowing in a transformer can be interrupted rapidly. A rather common device in some applications of dielectric heating is one for rapidly extinguishing arcs which can occur in intricate electrode systems. For example, in applications of sealing plastic materials together, such as in the manufacture of interior trim panels for automobiles, a rather intricate decorative pattern is embossed into the plastic material. The dies for forming this pattern are often complex and may require many man-hours to construct. In this application they are energized while under heavy pressure with the high-frequency high voltage to accomplish the sealing and embossing.

If the plastic material has a defect, such as a small metallic inclusion, or if for some reason a high-voltage transient occurs, an arc may form between the embossing electrodes which would damage them. To avoid this, protective devices are included which will very rapidly stop the oscillation in the generator. The stopping time ranges from microseconds to a few milliseconds but the current interruption occurs in several microseconds in any case, (i.e., di/dt is high). Very often these devices, by one means or another, apply high voltage to the grid of the oscillator tube. Experience shows that, regardless of the rate of build-up of this voltage, the oscillation dies out due to a characteristic inherent in the r-f generator and this decay is very rapid. Fig. 3 shows a typical oscillator circuit in which a high-speed vacuum switch has been placed in the grid circuit. Opening this switch causes oscillations to stop because of charging of condenser C_1. In an attempt to slow down the rate of cessation of oscillation, a series R-C (resistance-capacitance) network was placed across the vacuum switch.

Table I shows the results of varying the capacity of the R-C network shown in Fig. 3. The maximum and minimum E is the deflection of the oscilloscope in centimeters, but is proportional to the magnitude of the transient. The time figures show the interval from the instant the switch is energized until oscillation is stopped. Perusal of the table indicates that, regardless of the amount of the capacity in the suppression network, the value of peak transient voltage is essentially the same. Since the load on the oscillator was constant for these tests, this indicates that the rate of change of direct current at cutoff time is constant for all values of C. However, the time of operation of the protective system is greatly increased by increasing the suppression capacity. This effect is undesirable as the object of the device is to stop oscillations as rapidly as possible.

Thus, when an arc-stopping device is used, the current in the plate transformer is suddenly interrupted. A high-voltage transient is generated, the magnitude of which depends primarily on the value of the current which was interrupted. Curves have been taken which indicate this proportionality; see Fig. 4. The two lines indicate the maximum and the minimum transients read in a series of 100 tests for each point. The dashed-line curve, which is approximately the median line, was drawn through the origin to indicate the linearity of the curve.

Fig. 5(A) illustrates typical transients at the d-c output of the rectifier when the load current is interrupted by the arc-stopping device. The transients are very similar excepting the magnitude only, which is proportional to the current interrupted as indicated in Fig. 4.

The shape of the high-voltage transient generated usually does not have an extremely fast rise time, 30 μsec (microseconds) or so, and unless the interruption occurs at abnormally high currents such as might occur in an overload, the transient is not particularly damaging to a transformer. However, such devices often operate when an abnormal current is flowing and this can cause very high voltage to occur between individual turns in a given transformer. It should be remembered that since the current in a transformer is everywhere the same, the transient voltage is induced in each turn in accordance with the effective inductance of that turn. Thus it is entirely possible, because of the relatively slow rise time of this type of transient, to have excessive voltages occur between each turn throughout the entire winding of the transformer. As a result, and probably because of the usual practice of increasing the insulation of end turns, failure often occurs in turns near the center of the winding rather than near the terminals. There is a corollary effect of this type of transient, described in the following section, which can be exceedingly damaging of the transformer insulation.

In communications, similar high-speed devices are sometimes used to extinguish arcs which may occur in r-f transmission line systems due to voltages induced by lightning surges. It is probable that in such cases the d-c filter used to smooth the ripple of rectification sufficiently isolates the plate transformer from the rapid interruption of current.

Another cause of current interruption in a secondary of a rectifier transformer is known as "ion starvation." Hull and Elder[1] (and Steiner and Strecker in an unpublished paper) describe this phenomenon which can occur when a soft tube rectifier is called upon to pass more

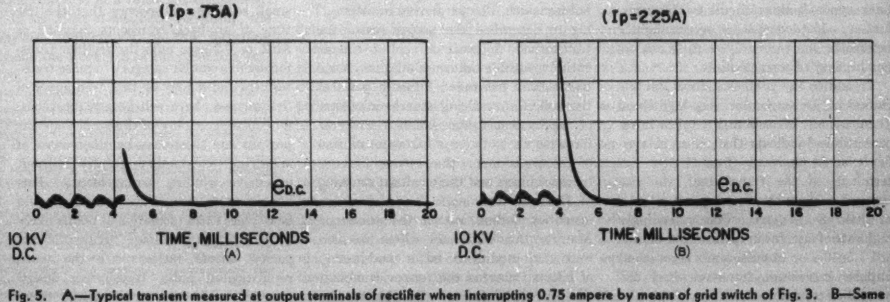


Fig. 5. A—Typical transient measured at output terminals of rectifier when interrupting 0.75 ampere by means of grid switch of Fig. 3. B—Same as (A) except current interrupted 2.25 amperes

current than there are ions available to carry the current. Normally, under rated temperature conditions, the ion starvation point of a rectifier is considerably higher than the short-circuit current of a transformer. Ion starvation cannot occur under these desirable conditions and perhaps this explains why it is a little known phenomenon. However, in industrial equipments, where less attention is given to rectifier temperature by the user of the equipment, it is quite possible to have ion starvation occur because of cold rectifier tubes.

Ion starvation occurs when magnetic forces due to heavy current flow cause the ions to leave momentarily the normal current path, so that the current is interrupted in not more than a few microseconds. After interruption, the ion starvation disappears and then ions are available for carrying current once more. If the short circuit is still present, as is usually the case, ion starvation may occur again. This may recur many times before the protective circuit-breaker equipment can operate. Thus, when ion starvation occurs, a transformer may receive many closely spaced interruptions of secondary current. This induces very high-speed very-high-voltage transients within the transformer windings.

Combination Insulation and Coil Movement Failures

Inasmuch as insulation failure is often caused by interruption of current and the coil movement failure by short-circuit currents, it would seem incompatible to have a combination failure. However, in rectifier services these two failures can occur simultaneously. When a high-voltage transient is caused as by interruption of current flowing from a rectifier, the high-voltage transient induced in the transformer is applied to the rectifier tubes, often causing an arc back. This arc back applies a short circuit to the transformer. The overvoltage occurs slightly before the short circuit, but both can occur because of a single fault.

To add to the problems, short circuits caused by arc backs are very-high-speed phenomena. Measurements taken in recent studies[2] indicate that, when a rectifier tube is physically located near the terminals of the transformer, the rise time of a short-circuit transient can be as little as 0.2 μsec. This exceedingly high rate of rise, far in excess of the standard 1.5x40-μsec impulse wave from usual impulse generators, can cause very abnormal distribution of voltages within the windings of a transformer.

As previously mentioned, the induced voltage transient, due to interruption of current in the secondary of the transformer, is quite evenly distributed throughout the winding because of its relatively slow rise time. In contrast to this, the rapid arc back of the rectifier tube effectively connects together the two ends of the transformer winding. Since this is equivalent to a steep wavefront negative impulse at the terminals of the transformer, the distribution of voltage due to this impulse is not uniform unless special precautions are taken in the design of the transformer.[2]

The studies made recently show that, in a high-voltage winding which was constructed in 16 parts with taps brought out to measure part voltages, the maximum voltage appeared across the first section and was equal to 36% of the total peak induced voltage across the whole coil. As this wave progressed through the transformer, the voltage across other portions was less, but still high. In several interior parts of the winding the voltage was as much as 25% of the total peak induced voltage. Thus, instead of the normal one sixteenth of the terminal voltage, this section received one fourth of the total terminal voltage which was four times its normal voltage.

Furthermore, since an arc back is most likely to occur when an abnormally high voltage is placed on the rectifier tube, as can be done when the load current on the transformer is suddenly interrupted, it is easily possible for the four-times-normal voltage to reach eight or twelve times normal. This much voltage applied across an interior section of a total winding which, in normal transformer design, is not usually insulated for high transient voltages, can easily destroy the insulation. This may explain why many failures occur in interior parts of the winding.

It is probably this combination type of failure which is so perplexing in attempting to determine the proper reason for transformer failures. It is, of course, entirely possible that once a liquid-cooled transformer has failed between parts of the coil, the resulting severe unbalance of mechanical forces causes the transformer coils to move. Likewise, motion of the transformer, due to many sledge hammer blows and the resultant crushing of the end turn insulation, may cause insulation failure within the windings. Many of the failures the writer has seen were attributed to the latter combination of effects, whereas the former combination was probably as responsible for the actual failure.

Some Solutions to the Problems

Solutions to the foregoing problems will be found in co-operative actions of both the transformer manufacturer and the equipment designer. There are contributions which each can make but neither can solve the problems alone.

In the case of coil movement, bracing of the transformer, as mentioned earlier is the best solution. This is primarily the responsibility of the transformer designer. The equipment designer can help by using a high-speed circuit breaker and by including some form of current limitation. Both have been used in equipments and have proved advantageous. Neither will replace adequate bracing of the transformer during its manufacture.

High-speed circuit breakers are limited as to how fast they will remove a short circuit. A modern circuit breaker is considered to be high-speed if it operates between 1 and 2 cycles. This is between 17 and 34 milliseconds. Oscillograms indicate that the short-circuit current in a transformer supplying a 3-phase rectifier will reach its maximum value well before one cycle. Therefore, while a high-speed circuit breaker cuts the duration of an applied short circuit, it cannot cut the intensity of the force.

Current limitation is useful, particularly for very stiff power lines. However, this adds regulation problems, usually minor but not always, and it also adds transient problems during current interruption. Likewise, the expense of current limitation devices is considerable and in excess of the cost of adequate bracing within the transformer.

High-speed high-voltage transients require considerable care in the design of the transformer to ensure that the distributed capacity of the winding(s) is such that even very high-speed transients will be distributed throughout the entire winding rather than piling up on small sections. Discovery that the rise time of arc-back transients may be as little as 0.2 μsec puts the rectifier transformer in a similar category to pulse transformers and it may be that techniques of design used there will be advantageous. Of course, economy of design is very important and there may be other ways of achieving satisfactory results without expensive winding arrangements. Perhaps a more economical solution would be to place extra insulation at points likely to receive extra voltage because of capacity effects, rather than the more sophisticated pulse transformer design techniques. It may prove equally effective in stopping the failures.

The equipment designer can help to increase the rise time of the arc-back transients by including a little inductance between the transformer terminals and the rectifier tubes. Spacing the leads widely, use of inductive resistors in the plates of the rectifier tubes, etc., will help to reduce the rise time. If the rise time can be kept to more reasonable values the transformer designer's problem is eased.

Several types of transient suppression devices are available. Thyrite or other nonlinear surge resistor suppressors may be connected across the terminals. In practical cases these may so reduce the transient that the terminal voltage does not exceed twice normal. *R-C* networks may also help, as will protective gaps. In many cases these devices are not adequate in high-voltage rectifiers because of the necessity of using the tubes closer to their maximum ratings than practical transient suppression will permit.

It is obvious that transient suppression means which are connected to the terminals of the transformer may not ensure adequate suppression of transients throughout the windings of the transformer. It is only the terminals of the transformer that are available to the equipment designer. The transformer designer may be able to include suppression to advantage, within the windings.

It is believed that adequate understanding of the various phenomena which may occur in a transformer applied to high-voltage rectifier service is essential to the best solution of the over-all design problems. These should be studied in detail by the designers both of transformers and of equipment. Co-operative efforts on the part of both are required to obtain the optimum in cost reduction without sacrificing reliability.

References

1. The Cause of High Voltage Surges in Rectifier Circuits, Albert W. Hull, Frank R. Elder. *Journal of Applied Physics*, New York, N. Y., vol. 13, June, 1942, pp. 373–77.

2. Transient Voltages in Rectifier Transformers, B. C. Biega, H. W. Lord. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 78, May 1959, pp. 83–90.

Design of a Vertical Magnetic Recording Head with Large-Scale Models

J. T. SMITH
NONMEMBER AIEE

Synopsis: A technique is described to aid in the design of magnetic recording heads. Large-scale models are used for the experimental optimization of recording head designs. To illustrate this technique, appropriate scaling factors are derived and then applied to the design of a vertical recording head.

IT IS WELL KNOWN that small-scale models of motors, transformers, and other electrical components are used to aid in design. These small models are advantageous in that certain parameters can be varied at a relatively low cost, when compared to the cost of varying the same function in a full-sized component.

In the field of magnetic recording, the physical dimensions of the recording heads, the thickness of the recording surfaces, and the gaps between the head and the surfaces are very small. For example, some of the ring-type heads have slots which are only a few microinches wide. These heads require many special techniques in their fabrication. As a direct consequence of this, to change one of the parameters in the recording head while holding all other variables constant is a difficult and expensive procedure. In this paper large-scale models of recording heads are used to relax the precision of machining required for studying the effect of several variables upon the resolution and amplitude of magnetically recorded signals.

The space covered by the magnetic field from a recording head is directly proportional to the resolution of the head. Hence it is desirable to have a linear increase of all lengths in a scaled model to produce a corresponding linear increase in the space covered by the magnetic field. Scaling factors are presented in the paper which will produce this point-by-point correspondence between the size of the magnetic field and the size of the head. These scaling factors are then applied to a vertical recording head having a particular structure. The technique is not limited to this structure but can be used upon any type of recording head. Emphasis is placed upon the ease with which the method may be used to obtain much useful information concerning the distribution of the magnetic fields around small objects.

The application of this method to a particular recording head is not done to show any advantages of this kind of recording over others; rather, it was felt that a successful application of the method presented in detail was the best way to show its usefulness.

A Scaling Experiment

Heuristic equations pertinent to large-scale models of vertical recording heads and surfaces have been developed in the Appendix. Approximate solutions to these equations have been evaluated and used to find the scaling factors shown in Table I. In every experiment using these equations, a certain degree of interpretation is required. For example, a simple experiment has been performed to illustrate the use and validity of these scaling factors. Two coils of 100 turns each were wound from no. *30* enameled copper wire (diameter of 0.01003 inch) upon coil forms whose dimensions and placements on a magnetic rod are shown in Fig. 1. A second rod and a second set of coils were made with all linear dimensions increased by a factor of 8 ($n=8$). In the second set, no. *12* enameled copper wire (diameter of 0.08081 inch) was used for windings.

An alternating current of varying magnitudes was passed through the primary coil, and the voltage developed in the secondary coil was recorded.

If B_t is the instantaneous field through the secondary coil, the secondary voltage is

---

Paper 59-148, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted May 2, 1958; made available for printing December 10, 1958.

J. T. Smith is with the International Business Machines Research Laboratories, San Jose, Calif.

Appreciation is expressed to Dr. E. A. Quade, to Dr. A. S. Hoagland, and to C. A. Bergfors for their many helpful suggestions. Through their initial efforts and continued encouragement this study was made possible. The author also acknowledges the helpful co-operation of J. J. Hagopian in furnishing the basic design and the models of the actual vertical recording heads.

Table I. Scaling Factors and Symbols

Parameter	Symbol	Scaling Equation	Comments
Length	l	$l_2=nl_1$	n = scaling factor
Cross-sectional area	A	$A_2=n^2A_1$	
Volume	l^3	$l^3=n^3l^3_1$	
Turns	N	$N_2=N_1$	but use a larger wire to occupy n^3 volume
Inductance	L	$L_2=nL_1$	
Pulse rise time	τ	$\tau_2=n^2\tau_1$	can be adjusted with the amplifier
Writing current	I	$I_2=nI_1$	for $B_2=B_1$
B-field	B	$B_2=B_1$	
Head-to-surface spacing	G	$G_2=nG_1$	
Surface velocity	v	$v_2=\frac{v_1}{n}$	
Surface thickness	T	$T_2=nT_1$	
Signal amplitude	e	$e_2=ne_1$	only for moving and recording heads
Signal width	λ	$\lambda_2=n\lambda_1$	
Thickness of undercoating	W	$W_2=nW_1$	if $W>>T$, this does not need to be adhered to: $\sim 10T$
Wire diameter	d	$d_2=nd_1$	
Permeability	μ	$\mu_1=\mu_2$	
Conductivity	σ	$\sigma_1=\sigma_2$	
Pulse frequency	ω	$\omega_2=\omega_1/n^2$	
Coil resistance	r	$r_2=\frac{r_1}{n}$	if the same material is used, this is already true

$$e_i=N_i\left(\frac{\partial\phi}{\partial_t}\right)=N_iA_i\left(\frac{\partial B_i}{\partial_t}\right)=N_iA_i\omega_iB_{0i}\cos(\omega_i t) \qquad (1)$$

where the subscript $i=1,2$ refers respectively to the small-sized or large-sized model. Consider the maximum secondary voltage as the quantity measured; thus, from equation 1

$$|e_i|=B_{0i}N_iA_i\omega_i \qquad (2)$$

Substitute the properly scaled parameters from Table I into equation 2 to determine the secondary voltage of the large-scale model in terms of the secondary voltage of the small-scale model. Note that the current in the primary coil of the large-scale system must be n times as large as the primary current in the small system.

$$|e_2|=B_{02}N_2A_2\omega_2\doteq B_{01}N_1n^2A_1\frac{\omega_1}{n^2}=B_{01}N_1A_1\omega_1=|e_1| \qquad (3)$$

Fig. 2 is a graph of the secondary voltages versus the primary currents where the primary current of the large scale model has been divided by 8. The coincidence of the curves is a partial check on the validity of the scaling factors.

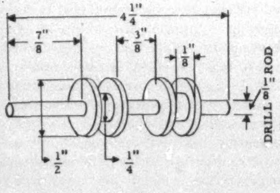


Fig. 1. Scaling factor experiment

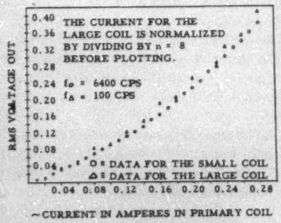


Fig. 2. Scaling experiment

Sensitivity Function

The distribution of the magnetic field around a recording head produced by a current in the windings is defined as the "sensitivity function," $H'(x, y, z)$; and the permanent magnetization left in the recording surface, as $M(x, y, z)$; see Fig. 3. A 2-dimensional plot of the sensitivity function and magnetization for a hypothetical vertical recording head is shown in Fig. 4.

If ϕ is the total flux from the magnetized area of film which threads the coils in the reading head, and v the relative velocity between the reading head and the magnetized surface, the signal voltage developed across the coil may be expressed by the following equation derived by Hoagland:[1]

$$e=K\frac{\partial\phi}{\partial t}=Kv\frac{\partial\left[\int_{-\infty}^{\infty}H'(x)M(x-x_0)dx\right]}{\partial x} \qquad (4)$$

where K is a constant determined by the surface thickness and its saturation distribution. From equation 4 it is apparent that the sensitivity function is used for weighting and represents the degree of coupling between the reading head and the magnetized information on the film.

In first approximation the magnetic recording surfaces used have a rectangular (square-loop) hysteresis loop, and as a consequence the changes in magnetization $[M(x)]$ left in the film are quite narrow when compared with the width of the sensitivity function. According to Hoagland,[1] if the magnetization is assumed to be a step function, i.e., to change from $-M_0$ to $+M_0$ at the point x_0, the voltage developed in the reading head is given by

$$e=2KNvM_0H'(x_0) \qquad (5)$$

where N is the number of turns on the reading coil. This equation states that the signal voltage developed in the reading head has the same shape as the sensitivity function.

Large-Scale Model (Enlarged 40 Times)

It is quite easy to obtain the sensitivity function for a large-scale head using a small pickup loop as a scanning device. If a properly scaled alternating current is passed through the writing coil of a 40×large-scale model head, an alternating H-field will be produced in the vicinity of

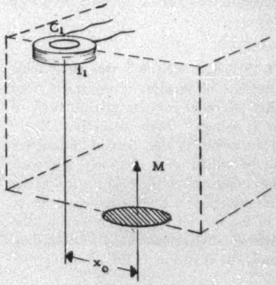


Fig. 3. Reciprocity

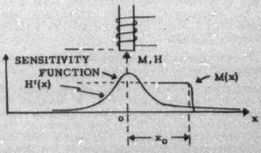


Fig. 4. Weighting

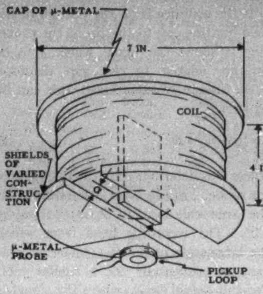


Fig. 5 (left). Simple vertical head

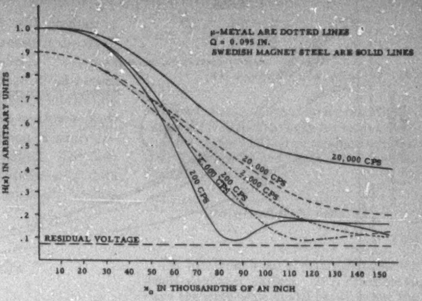


Fig. 6 (right). Sensitivity function versus frequency

the probe. A coil of wire whose diameter is small compared with the probe dimensions placed in this alternating field will have generated in it an electromotive force proportional to the maximum value of the H-field at that point. Thus a plot of these voltages as functions of distances from the probe will be equivalent to a distribution of the H-field around the probe. The orientation of the pickup loop is important since a voltage can be produced only if the component of flux threads it.

In the experimental measurements discussed, only that component of field perpendicular to the recording surface was measured, i.e., the axis of the pickup loop was perpendicular to the recording surface as shown in Fig. 5.

The magnetizing field $H(x)$ which produces the magnetization in the recording surface is much stronger than the reading field and could be much broader than $H'(x)$ because of saturation characteristics of the ferromagnetic metal in the probe of the recording head. In a large-scale model, if $H'(x)$ and $H(x)$ differ widely, then each must be handled by assuming $H'(x)=H(x)$ or by measuring an $H(x)$ for a large-writing current and another $H'(x)$ for an equivalent reading current.

A head enlarged 40 times was constructed as shown in Fig. 5. Because neither the wire size available nor the pulse amplitude were large enough to satisfy the conditions of scaling (i.e., $I_2=nI_1$ and $N_2=N_1$), it was decided to use more turns and a smaller wire size. In spite of this, good plots of the sensitivity function $H'(x)$ and the writing field pattern $H(x)$ were obtained since the rise times and coil resistance were not too critical at the frequencies used.

The pickup loop which was used to plot the distribution of the vertical component of the field from the probe of the large-scale vertical recording head had a diameter approximately one seventh of the width of the probe. A high-impedance vacuum-tube voltmeter was used to measure the voltage. The pickup loop was approximately 100 turns of 0.0025-inch-diameter wire, wound on an 0.012-inch-diameter core. The voltage induced in the pickup loop was in the order of 1 millivolt. Since mumetal is not readily available in thickness large enough to make probes 0.040 inch thick, the shields and probes of the large-scale recording head were laminated from 0.020-inch mumetal stock. To minimize electrostatic voltages being induced in the pickup loop, a common ground was made between the pickup loop and the current source (Hewlett Packard *305A* audio-oscillator). The electrostatic voltage and all other stray voltage not produced by the changing magnetic fields are indicated on the various graphs as "residual voltage."

Fig. 6 is a plot of the e, or $H(x)$ where $H(x)$ is the vertical component of H which changes with the spatial distance x_0, versus the distance x_0 away from the center of the vertical probe. This plot is made for two types of metal and at three different frequencies. Note that the width of $H(x)$ is dependent upon the type of metal

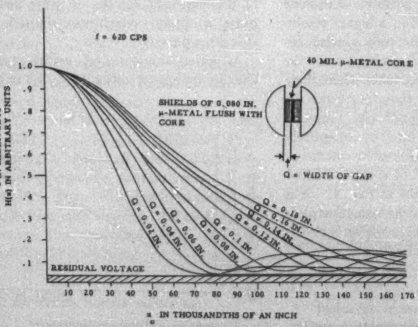


Fig. 7. Sensitivity function versus shield gap for aluminum-backed writing surface

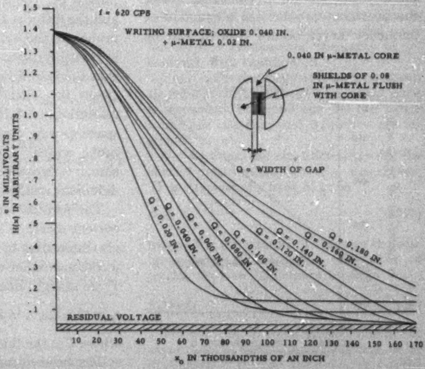


Fig. 8. Sensitivity function versus shield gap for mumetal-backed surface

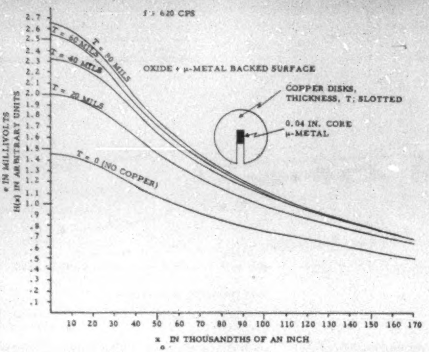


Fig. 9. Sensitivity function versus x, for eddy-current shielding

in the probe and also depends on the frequency. At 200 cycles the edge of the metal of the shields nearest the probe begins to have induced in it an appreciable magnetization of opposite polarity to that of the probe. This would result in an "overshoot" which is a reserval in the direction of magnetization left on the recording surface and causes a corresponding change in polarity of the output voltage induced in the head during reading. In general, this overshoot widens the signal width λ.

Although it is not shown, if the writing current in the large-scale head is reduced, the overshoot disappears for these probes. As illustrated in Fig. 6, the mumetal is not affected by frequency as much as the Swedish magnet steel. It should also be noted that the frequencies for an actual head corresponding to the 200-cps (cycle-per-second), 2,000-cps, and 20,000-cps curves are 320 kc, 3.2 megacycles, and 32 megacycles. For most recording heads, the overshoot problem as a function of frequency never occurs since it can be controlled by reducing the writing current. While the cause of these effects has not been explained, the effects have been found to exist in the small head.

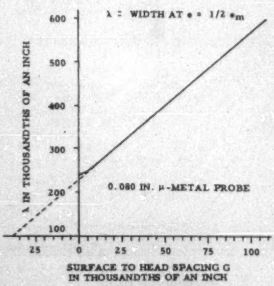


Fig. 10. Resolution as function of head-to-surface spacing

Figs. 7 and 8 show a plot of $H(x)$ versus x_0 with the shield gap Q as a variable. An aluminum-based writing surface is compared with a mumetal-based writing surface in these two figures. The writing surfaces were prepared by coating iron oxide particles (γ-Fe_2O_3) held in an organic binder upon the metal surfaces, and the 0.040-inch coating was deposited upon 0.020-inch metal. It is interesting to note that the width λ of $H(x)$ and $H'(x)$ at the base is linearly related to the width of the gap between the shield and the probe i.e.,

$$\lambda = bQ + C \tag{6}$$

where $b \cong 2$ for a mumetal-based writing surface and $b \cong 1.7$ for an aluminum-based writing surface. The constant C is determined by the thickness of the probe and the distance G, between the tip of the probe and the writing surface. Since one is normally interested in a higher resolution head (and because wide spacing between the shield and probe lowers the resolution) no data were taken where these spacings were greater than four times the probe width. Hoagland[2] has found a similar λ dependence on Q by an entirely different method.

The focusing effect of a slotted eddy-current shield is shown in Fig. 9. It is also interesting to note the logarithm plots of the maximum voltages of Fig. 9 versus T are straight lines indicating that

$$e_m = H(x_0)_{\max} \propto \ln KT \tag{7}$$

where T is the thickness of the shield.

The head-to-surface spacing is an important parameter which was studied with the 40×large-scale model. This spacing G is the distance from the bottom of the field plotting coil to the tip of the probe. Fig. 10 shows the signal width λ to be linearly related to the gap distance. With the use of equation 6, this can be written as

$$\lambda = bQ + mG + D \tag{8}$$

where D is determined by the thickness of the probe. In Fig. 10 m is determined for the large-scale head to be approximately 3.25. The recording head from which the large-scale model was patterned was used experimentally for a similar plot. For this actual head (where the signal is rewritten at each gap) m was found to be 3.3, which is in excellent agreement with the large-scale head.

Large-Scale Model Enlarged Five Times

In the 40×head the sensitivity function, $H'(x)$, and the writing $H(x)$, were found by measuring the voltage induced in a small coil of wire placed in the varying magnetic field around the probe. Because of its large size it was not feasible to use the 40×head to record on a moving surface. For this reason, a 5×head was constructed which was then used to write directly upon a scaled moving surface.

Equation 5 predicts a signal amplitude directly proportional to the velocity of the moving surface. This is experimentally verified in Fig. 11. However, equation 25 predicts the signal from the 5×head to be five times the signal amplitude from the 1×head, if the surface velocities are properly scaled. The signal from the 1× head at 1,500 inches per second is 45 millivolts, and the signal from the 5× head at 300 inches per second is 15 millivolts; this is a ratio of 3 to 1. Neither the surface to head spacing G nor the thickness of the magnetic coating W were scaled quite properly, which contributed to this discrepancy.

As stated previously, the width of the changes in magnetization $\lambda_{M(x)}$, left in the

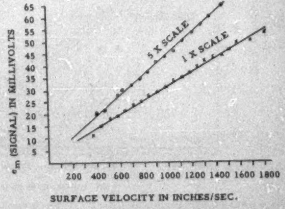


Fig. 11. Surface velocity versus signal voltage

Table II. Large (5X) Dynamic Head Data for Magnetization Width Studies

Units	5× Write* e_s, Millivolts	5× Write* Signal Width, λ, Inches	1× Write† e_s, Millivolts	1× Write† Signal Width, Inches
5× read	42	0.043	2.6	0.036
1× read	8.8	0.022	23	0.0076

* $w=0.0035$ inch; $i_w=0.57$ ampere.
† $w=0.0009$ inch; $i_w=0.2$ ampere.

magnetic surface, should be narrower than the width of the sensitivity function $\lambda_{H'(x)}$. This is expected because in the first approximation the magnetic recording surfaces have square hysteresis loops and the writing field $H(x)$ must reach the effective coercive force of the surface material before the magnetization in the surface changes.

From equation 4 it may be seen that the maximum recovered signal width λ_e cannot be greater than the width of the sensitivity function $\lambda_{H'(x)}$, added to the width of the magnetization $\lambda_{M(x)}$:

$$\lambda_e \leq \lambda_{H'(x)}+\lambda_{M(x)} \tag{9}$$

To obtain some information about the relative contributions of writing and reading to the base width of the recovered signal (signal width on readback), a magnetic bit was recorded with the small head (1×) and read by both the large (5×) and the small head; a bit was recorded also by the large head and read by both the large and the small head. The data are given in Table II.

From Figs. 8 and 9, the $\lambda_{H'(x)}$ for both the 5× and 1× heads can be calculated to be about 0.030 inch and 0.006 inch respectively. Using Table II and equation 9, $\lambda_{M(x)}$ for both the 5× and 1× heads can be determined to be 0.013 inch and 0.0016 inch respectively. The recovered signal width for writing with the 1× head and reading with the 5× head should be given by $\lambda_{M1(x)}+\lambda_{H2(x)}\cong 0.032$ inch, which compares favorably with the measured value of 0.036 inch. Similarly, the recovered signal width for writing with the 5× head and reading with the 1× head should be $\lambda_{M2(x)}+\lambda_{H1(x)}\cong 0.019$ inch, compared to the measured value of 0.022 inch. These qualitative measurements show that for a given recording head used both for writing and reading, the largest portion of the recovered signal width is contributed by reading.

Conclusions

Scaling factors appropriate to magnetic recording heads have been derived in such a way that a linear increase in all lengths in a large-scale model will produce a corresponding linear increase in the space covered by the magnetic field.

The principal advantage of the technique is to relax the precision of fabrication and measuring tolerances so that recording resolution as a function of many parameters may be easily studied. The close agreement between the experimental data from the actual and scaled heads justifies the technique.

Appendix

Theory

In this Appendix some of the equations pertinent to large-scale models of recording heads and systems will be developed. It should be emphasized that, even though the equations and their solutions are not exact, the analogue method used should yield useful information.

D-C Field

The reluctance of a path i in a magnetic recording head is

$$R_i=K\frac{l_i}{\mu_i A_i} \tag{10}$$

where l_i is the length, A_i is the cross-sectional area, and μ_i is the permeability of the path.

Assume a linear increase of n for all dimensions of the path so that $l_{i2}=nl_{i1}$ and $A_{i2}=n^2A_{i1}$; where the second subscript specifies the head being discussed, i.e., the subscript 2 refers to the large-scale head. Thus the ratio of reluctance for corresponding paths in the two heads is

$$R_{i2}=\frac{R_{i1}}{n} \tag{11}$$

provided $\mu_{i2}=\mu_{i1}$

A-C Field

The attenuation of a varying B-field by eddy currents in a semi-infinite slab can be shown[8] to be

$$B_i=B_{0i}\exp\left[-(1+j)\sqrt{\frac{\omega_i\mu_i\sigma_i}{2}}x_i\right] \tag{12}$$

where

B_{0i} = initial B-field at surface
ω_i = frequency of variation of incident B_{0i} field
μ_i = permeability of semi-infinite slab
σ_i = conductivity of semi-infinite slab
x_i = depth of penetration

A heuristic approach is used where the B-field reaching any particular point is used to calculate a pseudoreluctance R_i, i.e.,

$$R_i=\frac{\text{mmf}_i}{\phi_i}=\frac{\text{mmf}_i}{B_iA_i} \tag{13}$$

Thus, with the use of a linear expansion of n times for all dimensions as before, the ratio of reluctance of varying B-fields for any path i becomes

$$R_{i2}=\frac{R_{i1}}{n}\exp\left\{\frac{(1+j)x_{i1}}{\sqrt{2}}\times\left[n\sqrt{\omega_{i2}\mu_{i2}\sigma_{i2}}-\sqrt{\omega_{i1}\mu_{i1}\sigma_{i1}}\right]\right\} \tag{14}$$

where reluctance R is a complex quantity.

Since it is desired to keep the same reluctance for every path (air or conducting material) equations 11 and 14 must be identical. For this to be true

$$\frac{(1+j)x_{i1}}{2}\left[n\sqrt{\omega_{i2}\mu_{i2}\sigma_{i2}}-\sqrt{\omega_{i1}\mu_{i1}\sigma_{i1}}\right]=0 \tag{15}$$

The hysteresis loops of materials make the heads nonlinear devices. It is desirable to have the large-scale model heads follow the same nonlinear response curves as the small head. If the H field at every corresponding point in space and time for both heads is the same, and the identical construction material is used in each head, then $\mu_{i2}=\mu_{i1}$ and $\sigma_{i2}=\sigma_{i1}$, and the heads will have the same nonlinear response. For a linear increase in all dimensions of a current-carrying circuit, it may be shown from Ampere's law that $H_{i1}=H_{i2}$ if one assumes $I_{i2}=nI_{i1}$.

From the preceding assumptions combined with equation 15 the scaling condition imposed upon the frequencies becomes

$$\omega_{i2}=\frac{\omega_{i1}}{n^2} \tag{16}$$

The previous theory applies to the distribution of the magnetic field around the writing probe and in the magnetic recording surface. The distribution of the magnetic field around the writing head is defined as the sensitivity function, $H(x, y, z)$, and the permanent magnetization left in the recording surface as $M(x, y, z)$.

Dynamic Head

In the preceding section only the sensitivity function was considered. To complete the picture the remanent magnetization in a large-scale moving surface must be considered.

First, consider the writing pulse of current; every harmonic of the fundamental frequency in the original pulse must exist in the same proportion in the expanded pulse for the new fundamental frequency in accordance with equation 16.

From a Fourier analysis it may be shown that this condition exists if the current pulse is expanded in amplitude by the factor n and the rise time and frequency are expanded by n^2.

To maintain the condition $H_{i2}=H_{i1}$ the current relationship, $I_{i2}=nI_{i1}$, was chosen. This automatically fixed N_2 to be

$$N_2=N_1 \tag{17}$$

where N is the number of turns on the coil. Since inductance has the dimensions of length, the new inductance becomes

$$L_2=nL_1 \tag{18}$$

For equation 17 to be satisfied and at the same time to allow a linear increase in all dimensions of the coil, the diameter d of the wire should be

$$d_2=nd_1 \tag{19}$$

The time constant of the inductance of the head is a function of the resistance r of the coil as well as the inductance; hence

$$r_2=\frac{r_1}{n} \qquad (20)$$

Since the bits must be spread out on the new large-scale moving surface exactly n times as far apart, the surface velocity becomes

$$v_2=\lambda_2\omega_2=(n\lambda_1)\left(\frac{\omega_1}{n^2}\right)=\frac{\lambda_1\omega_1}{n}=\frac{v_1}{n} \qquad (21)$$

The signal amplitude e is given by

$$e_i=N_i\left(\frac{\partial\phi_i}{\partial_t}\right)=N_i\left(\frac{\partial\phi_i}{\partial x_i}\right)v_i \qquad (22)$$

where N is the number of turns in the recording head and ϕ is given by

$$\phi=SK\int_{-\infty}^{\infty}H(x)M(x)dx \qquad (23)$$

where $H(x)$ is the sensitivity function, $M(x)$ the recording surface remanant magnetization, S the effective cross-sectional area of $M(x)$, and K is a constant determined by the surface characteristics.

Since $S_2=n^2S_1$, $x_2=nx_1$, and $H(x_2)=H(x_1)$, and assuming $M(x_2)=M(x_1)$, it follows that

$$\left(\frac{\partial\phi_2}{\partial x_2}\right)=\frac{n^2S_1}{k}\,\frac{\partial}{\partial x_2}\left\{\int_{-\infty}^{\infty}H(x_1)M(x_1)dx_2\right\}=n^2\left(\frac{\partial\phi_1}{\partial x_1}\right) \qquad (24)$$

and

$$e_2=N_2\left(\frac{\partial\phi_2}{\partial x_2}\right)v_2=N_1\left(n^2\frac{\partial\phi_1}{\partial x_1}\right)\frac{v_1}{n}=ne_1 \qquad (25)$$

A summary of the scaling factors is given in Table I.

References

1. Magnetic Data Recording Theory: Head Design, **Albert S. Hoagland.** *AIEE Transactions*, vol. 75, pt. I (*Communication and Electronics*), Nov. 1956, pp. 506–12.

2. High Resolution Magnetic Recording Structures, **A. S. Hoagland.** *Journal of Research and Development*, New York, N. Y., vol. 2, Apr. 1958, pp. 90–104.

3. The Electromagnetic Field in Its Engineering Aspects (book), **G. W. Carter.** Longmans, Green and Company, Inc., New York, N. Y., 1954, p. 245.

4. Studies of Magnetic Recording, Part II, **W. K. Westmijze.** *Philips Research Reports*, Eindhoven, Holland, vol. 8, June 1953, pp. 161–83.

Properties of Modern Teleprinters with Regard to Signal Transmission

HERBERT WUSTENEY
NONMEMBER AIEE

Synopsis: The properties of a teleprinter with regard to signal transmission are determined by the performance of three functional units: the drive system, the transmitter, and the receiver. Their design and characteristics in the pageprinter recently developed by Siemens & Halske are subjected to a theoretical examination and practical results discussed. Results coming up very closely to the ideal values have been achieved with simple means.

TODAY the large majority of all teleprinters are installed in the premises of private customers and not in the telegraph offices. These machines are intended for direct record communication between the various offices of the same or of different companies or governmental agencies, and their place is at the side of the office typewriter. In the future, the control of manufacturing processes, the interoperation with computers and with similar equipment for transmission of data will assume, in addition to the present exchange of messages, a very important role.

A modern teleprinter must be tailored to these applications. Therefore it should resemble a conventional office typewriter as regards its outward appearance, it should display elegant styling, and it should be easy to operate. A low operating noise level is highly desirable. Furthermore this machine should work satisfactorily when operated by untrained people and the necessity for servicing by a trained maintenance mechanic should be limited to the bare minimum. Quite a number of companion units are required to meet the various operational requirements and installation as well as removal of these attachments should not present undue difficulties.

If the same machine is supposed to satisfy also the requirement imposed by military organizations for satisfactory operation under severe conditions, operation on board of ships or airplanes for instance, these requirements are likely to have a significant effect on the over-all design of the equipment.

The Teleprinter Is a Telegraph Apparatus

Under all applications the teleprinter is, however, a telegraph apparatus which, to an ever-increasing extent, is required to ensure optimum reliability in message transmission. As a result, great importance must be attached to its properties in this respect. Most of the present teleprinter circuits are made up of a number of line sections with interposed switching centers and carrier systems. Considering the tremendous number of existing teleprinter channels, the use of regenerative repeaters cannot be justified from an economical point of view. Moreover, doubts are already being raised as regards the advisability of employing regenerative repeaters because it is safe to assume that teleprinter channels will some time in the future be used for the transmission of data by means of codes having six, seven, eight and more significant elements which cannot be handled by 5-unit code regenerative repeaters.

Siemens & Halske have recently started production of a new pageprinter model (Fig. 1) which has been designed to meet the afore-mentioned requirements. Properties with regard to signal transmission have been one of the prime considerations.

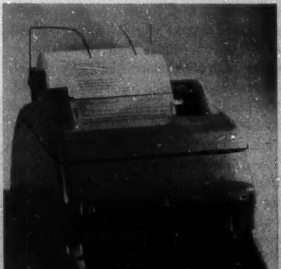


Fig. 1. Siemens pageprinter 100

Paper 59-154, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 15, 1958; made available for printing December 3, 1958.

Herbert Wusteney is with Siemens & Halske AG, Munich, Germany.

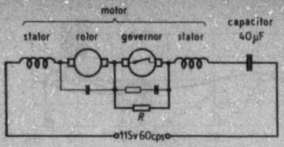


Fig. 2. Circuitry of governed motor

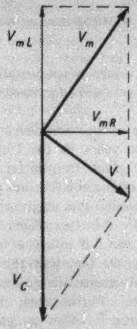


Fig. 3. Voltage values during start-up period at 500 rpm

V = 115 volts, mains voltage

V_m = 162 volts, voltage across motor terminals

V_c = 201 volts, capacitor voltage

The latter are decisively influenced by the following three factors:

Speed of the drive unit with its effect on transmitter and receiver.
Outgoing signal distortion.
Receive margin of selector mechanism.

Motor Speed

COMMUTATOR MOTORS

Many teleprinters are powered by commutator motors with centrifugal contact governors. With a constant load and a constant supply voltage it is easy to achieve a constant speed for a certain period of time. Teleprinters, however, are required to tolerate considerable voltage fluctuations. The load imposed on the drive unit varies during every single operating cycle and, in addition, it depends on the temperature range found under service conditions. An additional requirement is that the start-up time, i.e., the time passing before rated speed and operational readiness are reached, must not exceed one second.

As is known, constant speed in spite of load fluctuations can be attained by increasing the inertia of the rotating masses. To take advantage of this, without excessively prolonging the start-up time, the Siemens pageprinter has a capacitor connected in series with the motor; see Fig. 2. Thus, for instance, a 40-μf (microfarad) capacitor is employed with motors rated for 115 volts, 60 cps (cycles per second). In the rest condition and at low speeds the motor acts as an inductance coil connected in series with the capacitor; see Fig. 3. As a result, the voltage at the motor terminals during the start-up time will be higher than the supply voltage. The effect is a higher start-up torque and reduced start-up time.

The motor has been rated for 3,750 rpm. Its speed has been purposely chosen higher by a certain amount than the synchronous speed of 3,600 rpm at 60 cps and 3,000 rpm at 50 cps, respectively. Governed motors which are operated at the synchronous speed tend to run at the synchronous speed for some time and to suddenly jump by one or by a number of cycles.

Even at an accurate average speed there is still the possibility of brief temporary fluctuations which might have an adverse effect. An electronic measuring method has therefore been employed to determine the actual speed at any given moment. Fig. 4 shows the results of this measurement at a speed of 100 wpm (words per minute).

Fig. 4 (right). Speed curve of governed motor 3,750 min^{-1} 100 wpm 220 volts 50 cps

a = 100 cps
b = line current
c = motor current
d = mechanical load (torque on main shaft)
e = phase position
$\Delta\omega_{max} = -1.35\%$
$\Delta\varphi$ = maximum amplitude of fluctuations during one signal = 0.22 millisecond $\triangleq$ 4.9 degrees $\triangleq$ 1.6% maximum distortion

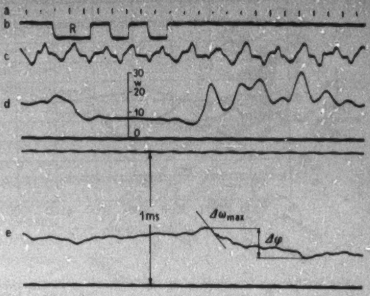
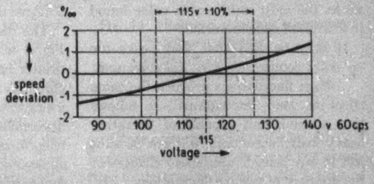


Fig. 5 (right). Influence of voltage on speed



The following characteristics are plotted:

1. 100 cps as the time base.
2. Line current (character *R* transmitted once).
3. Motor current.
4. Mechanical load on drive unit occurring during reception and printing of character *R*. These values have been determined with the aid of strain gages on the main shaft.
5. Phase position of motor shaft with relation to a crystal-controlled standard frequency of 1,000 cps. The bend in this characteristic represents the deviation from the nominal speed. It can be readily seen that the increased mechanical load (curve *d*) results in a momentary decrease in the motor speed. The total phase displacement during the time it takes to transmit one signal is not higher than 0.2 millisecond and this implies an increase in outgoing signal distortion of about 1.6% and a decrease of the receive margin by approximately 1.3%, respectively.

The effect of load fluctuations depends on their phase relation with respect to the operation of the governor (one contact

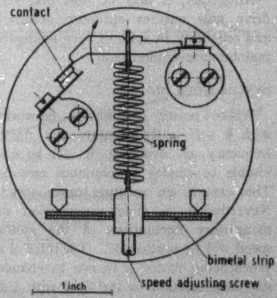
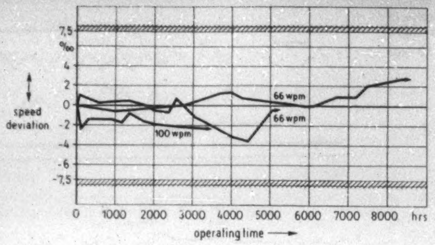


Fig. 6. Governor



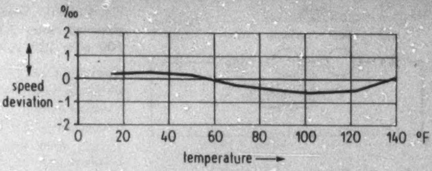


Fig. 7 (left). Influence of operating time on speed

Fig. 8 (above). Influence of temperature on speed

make with each revolution of the motor) and the current wave. The diagram shows the most unfavorable value found in a series of measurements.

It may be said that load peaks are virtually compensated by the drive unit.

Voltage fluctuations produce the same effect as load fluctuations. It follows that a drive unit capable of tolerating heavy load fluctuations will not be subject to any considerable speed changes in the presence of voltage fluctuations. Fig. 5 shows that this assumption holds true over an extended range.

Another factor tending to affect constancy of speed is contact erosion. It decreases the weight of the contact arm and the latter will come closer to the axle (Fig. 6). The governor spring has been designed to compensate this alteration so that readjustment of the speed will not be necessary for an extended period of operation (Fig. 7). The service life of governor contacts by far exceeds 5,000 operating hours.

Since the components of the governor, the steel spring in particular, are rather sensitive to temperature changes, a bimetal strip has been provided as a compensating means (Fig. 8).

The service life of the motor carbon brushes comes up to 3,000 operating hours.

Altogether, it may be said that this drive unit ensures optimum accuracy and reliability in spite of very moderate maintenance requirements.

SYNCHRONOUS MOTORS

Where power distribution networks with a sufficiently accurate and stable frequency are available, it will be advisable to employ synchronous motors. This ensures an accurate mean speed. Nevertheless, a close examination of momentary fluctuations of the motor speed will be worthwhile. The rotor of a synchronous motor is known to change its phase relation to the field with the load. It will lag with a higher load and lead with a smaller load. These changes in the phase position may affect outgoing signal distortion and receive margin to a degree that must not be underestimated.

The diagram in Fig. 9 corresponds to Fig. 4 with the exception that the teleprinter in question is powered by a synchronous motor. Phase displacements as a function of load and torque are clearly discernible and, during the length of one teleprinter signal, produce a shift by approximately 0.29 millisecond, corresponding to 2.2% start-stop distortion.

To avoid a still higher distortion, the synchronous motor had to be designed for a much higher power than was actually required for quick take-up and synchronous operation. In spite of this, the results obtained are less favorable than those obtained with a governed motor.

The pulsating torque is a characteristic of each synchronous motor. It produces gear noise. This can partly be offset by increasing the inertia. Nevertheless, gear noise will always exceed that of a pageprinter with governed motor.

Outgoing Signal Distortion

Distortion caused by speed fluctuations is augmented by the inherent distortion of the transmitter. It has been possible for quite some time already to keep this distortion very low. In particular, transmitters with six contacts, showed excellent results when installed in teleprinters designed for neutral current operation.

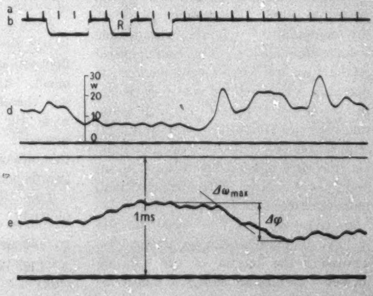
For the teleprinters developed during the recent years in the United States a single-contact transmitter has been adopted and the new Siemens pageprinter also has only one single contact. In addition to a further slight increase in accuracy, the advantage of this design consists in the fact that this send contact can be encapsulated to suppress radio interference. Furthermore, it can be easily converted into a polar-current transmitter. Siemens & Halske decided to adopt this design when, by means of a novel and simple arrangement, they succeeded in achieving excellent performance together with an appreciable reduction in the number of adjustments required. Fig. 10 shows the signal generator mechanism.

In the same manner as in the case of other well-known transmitter designs, the bifurcated rocker is shifted either to the right or to the left to prepare transmission either of a marking or of a spacing pulse. Each lobe of the transmitter cam raises the cam follower with its bifurcated rocker and the latter hits against the one or the other extension of the contact rocker bail. Rigidly connected with this bail is the contact plate which is made of insulating material. The control spring of the contact spring set rests

Vol.
Pt. 1
195

Fig. 9 (left). Speed curve of synchronous motor 3,000 min^{-1} 100 wpm 220 volts 50 cps

a = 100 cps
b = line current
d = mechanical load (torque on main shaft)
e = phase position
$\Delta\omega_{max} = -1.0\%$
$\Delta\varphi$ = maximum amplitude of fluctuations during one signal = 0.29 millisecond $\triangle$5.2 degrees $\underline{\triangle}$2.2% maximum distortion



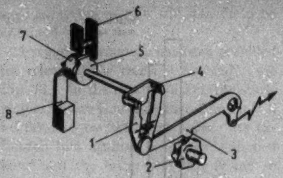


Fig. 10. Signal generator
1—Bifurcated rocker
2—Transmitter cam
3—Cam follower
4—Contact rocker bail
5—Contact control plate
6—Send contact
7—Friction brake
8—Leaf spring

against one of the two lugs of this contact control plate and the contact will be either closed or opened. When the cam follower with the bifurcated rocker is returned to prepare the setting for the next following signal element, the contact must be held in the position determined by the old setting. This is accomplished by a leaf spring forcing the friction brake against the contact control plate. The friction exerted by the spring and the brake holds the contact closed or open, as the case may be. This brake does not require any adjustment nor will it obstruct the movements controlled by the cam and it avoids bounce. Contact bounce proper has been practically eliminated by painstaking calculation and rating of forces, moving masses, and switching speed.

With this design of the transmitter and the measures taken to achieve accurate speed control it became possible to reduce the over-all outgoing signal distortion to less than 3% leading and lagging at a speed of 60 and 66 wpm and to less than 6% at 100 wpm even in the presence of mains voltage fluctuations amounting to ±20%.

A 10,000-hour endurance test carried out in our laboratories revealed that the distortion did not increase by more than 1/2% per 1,000 operating hours.

Receive Margin

The individual operations in a receiver designed for neutral current operation differ considerably from those of a polar-current receiver. Since the large majority of all teleprinters operates on neutral current, the following explanations shall be restricted to this operating option.

In all teleprinters now on the market the receiver has only one magnet and this magnet is called upon to control both the starting of the receiver shaft and the sensing of the five intelligence pulses. A closer study of the problem reveals, however, that these are two rather different functions which have very little in common.

Starting the camshaft of the receiver requires a motion which releases a previously arrested part so that the camshaft is free to revolve. Accurate timing and an excess of power are required. Sensing of the five intelligence pulses is effected merely by testing the position of the armature at certain instants.

To accomplish either of these two functions with optimum accuracy, the new Siemens pageprinter has been provided with a separate armature for each function and these armatures are controlled by a common selector magnet. Each of the two armatures is offered to the magnet by means of a motor-driven cam.

The following is a description of the pulse sampling process; see Fig. 11. The extension of the selector armature rides on the selector cam of the receiver camshaft. For each individual pulse, the selector cam offers the armature to the magnet, i.e., the selector armature is raised so far that only a small gap remains between armature and pole plates. At a closely defined instant, the armature is free to drop unless it will be held by line current flowing through the magnet. A few milliseconds after this release of the armature, the extension of the selector lever moves in the direction of the armature to determine whether the armature is held or released. The extension of the selector lever will continue to move below the armature if the latter is held by the magnet but this movement will be obstructed if the armature is released. Thus, the armature will assume one of two possible positions and, by way of intermediate members, this characteristic setting of the selector lever will be transferred to the code bars of the receiver.

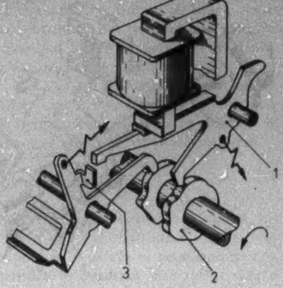


Fig. 11. Receiver selecting mechanism
1—Selector armature
2—Selector cam
3—Selector lever extension

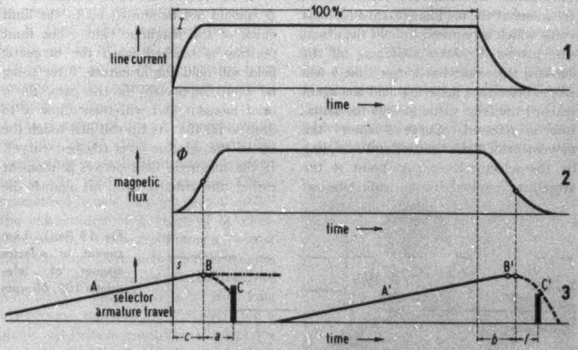


Fig. 12. Sampling limits of selector mechanism

A-A'—Offering of selector armature
B-B'—Release of selector armature
C-C'—Sampling instant
a—Sampling time
b—Decay time of magnetic field
c—Rise time of magnetic field
f—Dropping time of selector armature

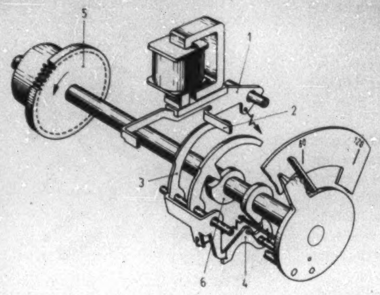


Fig. 13 (left). Receiver start-stop mechanism

1—Start-stop armature
2—Armature extension
3—Crescent-shaped lever
4—Locking lever
5—Friction clutch
6—Spiral cam disk

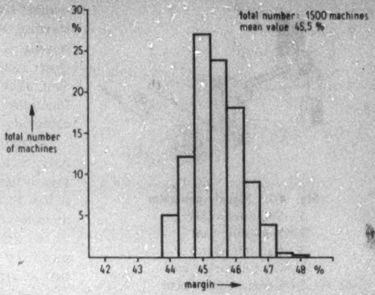


Fig. 14. Teleprinter 100 margin of receiver, 66 wpm

This sampling process takes place approximately in the center of the intelligence pulse. With an ideal receiver it should be possible to shorten each pulse at either side by 50% before a wrong signal is printed or, to express it in other terms, it should be possible to shift a pulse of exactly 100% length in either sense by 50% with regard to the sampling point. To achieve this postulate, certain timing conditions would have to be met which are shown in Fig. 12 for the receiver in question.

Curve *1* represents a current pulse with a contact make time of 100% exactly. The sloped wavefront is due to the self induction of the magnet while the trailing edge is elongated by a spark suppressing capacitor. The magnetic field is shown in curve *2*. The time *c* elapses between the moment the current is switched on and the moment the field has reached its limit value which is required to hold the armature attracted. After switching off the current, the somewhat longer time *b* will elapse before the decaying field has again reached the limit value so that the armature is released. Curve *3* shows the movement of the armature and sampling by the selector lever. At point *A* the armature is raised by the cam lobe, at instant *B* it is being released and allowed to drop unless it is held attracted. At instant *C* the selector lever is just moving into the range of the armature. The distance between *B* and *C* represents the sampling time *a*. Shown at the left-hand side (*A*, *B*, *C*) is one limit position and at the right-hand side (*A'*, *B'*, *C'*) the other limit position of the sampling process.

At the beginning of a current pulse the current may be so far displaced with respect to the sampling instant that instant *B* just coincides with the limit value of the magnetic field. If *B* is ahead of this point, the armature will drop and the selector lever will abut against the front edge of the armature (dashed curve). If *B* is behind this point, the armature remains attracted and the selector lever will not touch it (dash-dot curve).

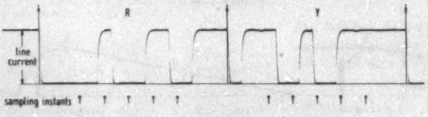
At the end of the current pulse, point *B'* should not be shifted up to the limit value of the magnetic field. The limit position is reached when the magnetic field will hold the armature, after being no more supported by the cam, for a brief moment and will then allow it to drop so far that its tip will just touch the tip of the selector lever (dashed curve). If the magnetic field decays a moment earlier, the armature will hit against the selector lever at *C'*. If the magnetic field will hold the armature a moment longer, the selector lever will pass below the armature.

If the displacement of sampling instant *C* between these two limits is to equal the ideal value of 100%, the following formula will be established with the aid of the diagram

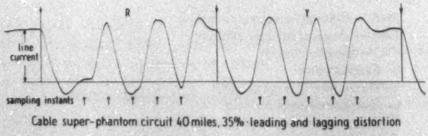
$$a = b - c + f$$

where *f* is the time it takes the armature to drop so far that it will touch the selector lever. The receiver of the Siemens page-printer has been designed to correspond to this formula.

The diagram shows that, during the sampling process, free movement of the armature is limited to only a very small distance so that the possibility of armature bounce is effectively eliminated. The armature is applied to the magnet very slowly and there is plenty of time for it to come to rest. Thus, the sampling process proper is in no way affected by the preceding armature movement or by the shape of the wavefront. It is of prime importance that the essential factors are determined by the design characteristics and are not dependent upon more or less accurate adjustments. Thus, if a disassembled receiver is being reassembled,



Off-line operation, 44% leading and lagging distortion



Cable super-phantom circuit 40 miles, 35% leading and lagging distortion

Fig. 15 (left). Line current in selector magnet of teleprinter 100, 66 wpm

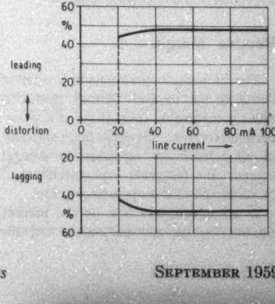


Fig. 16 (right). Teleprinter 100, margin, 66 wpm

Vol. 7
Pt. 1
1959

a wide receive margin will ipso facto be obtained.

A separate armature has been provided for releasing the receiver camshaft; see Fig. 13. In the rest condition, the start-stop armature is held attracted and it will be released by the start pulse. Its extension hits against the crescent-shaped lever and the latter will be pivoted so that locking lever is disengaged. The camshaft will now be turned by way of friction clutch. Since the start-stop armature is required again only at the end of the cycle to stop the shaft, it is raised by the spiral cam disk. It was possible to adopt a very long travel of the start-stop armature. It amounts to 15/32 inches, measured at the tip of the armature. As a result, abundant energy is available for release.

In the rest condition, there is a gap between the extension and the crescent-shaped lever. The armature is therefore allowed to drop freely and it has a considerable momentum when it hits against the crescent-shaped lever. This "striking release" proved to be very reliable.

When the teleprinters are being assembled, the receivers are not adjusted individually for optimum receive margin. Differences of margin resulting from manufacturing tolerances are disregarded. Fig. 14 shows the values measured on 1,500 pageprinters before their shipment to the customer. The mean value of the receive margin was found to be at 45.5% and the minimum value was 44%. These measurements have been carried out with a high-precision electronic distorted signal generator at 66 wpm.

This receiver does not only permit changes in the length of square-wave current pulses but tolerates also extreme distortion values in the case of sloped pulse shapes; see Fig. 15.

Variations of the line current will hardly affect the receive margin (Fig. 16), and it is immaterial whether the machine is operated at, say, 40 or 60 milliamperes.

Conclusions

By selecting a suitable mechanism, by careful rating of components, and by close attention to details in manufacture it is possible to construct with rather simple mechanical means a teleprinter which comes very close to the theoretical limit values as far as outgoing signal distortion and receive margin are concerned.

More About Nonarmored Submarine Cable

C. S. LAWTON
MEMBER AIEE

IN JANUARY 1953 a report[1] was made of the design, manufacture (1950), and laying (October 3, 1951) of an experimental length of novel nonarmored deep-sea submarine telegraph cable[2,3] in 1,250 fathoms of water. In the course of the discussion of the paper it was stated that after a test of about 3 years it was hoped to recover the cable and present a further report on its condition.

Cable Recovery

Actually no suitable opportunity was found to pick it up before the old coastwise section in which it had been installed was taken out of service in January 1955 with the closing of Western Union's Canso, Nova Scotia, station. In July 1955 while the *C. S. Cyrus Field* was in the vicinity an attempt to recover it was made but the ship's dynamometer was insensitive at the low tensions required and the speed of picking up was excessive. The cable parted near the splice to the transitional type on one end, and the bight splice to the armored cable on the other end. No further ship time could be spared until last summer when the *C. S. Lord Kelvin* recovered the entire length laid in 1951 with the exception of 150 fathoms pulled off the ends and recovered in the previous attempt along with the transitional types.

The recovery of this experimental cable on July 26–27, 1958, is of particular significance in the light of recent developments. In 1956 the British Post Office announced "a new deep-sea coaxial cable" the culmination of 5 years of development work,[4,5,6] thus confirming a forecast made in the afore-mentioned discussion that "No obstacle is seen to the design of nonarmored coaxial cable." Not long after came news of the proposal for a single All-British transatlantic telephone cable (to handle 60 simultaneous 2-way telephone conversations) which would be nonarmored throughout the deep-water portion of the route.[7] In July 1958 it was announced that this cable definitely would be a reality by 1961. As a matter of fact, signaling frequencies in carrier-operated transoceanic telephone cables rapidly are approaching the realm where steel-armor wires must be eliminated if attenuation is to be kept within bounds. Whereas laying and recovery tests have been made with nonarmored coaxial cable of British design, there have been no actual service tests to our knowledge.

Both ends having become detached, recovery of the experimental cable with which this report deals had to be effected by grappling for the small cable itself. It was hooked on the second tow 2.04 nautical miles from the east end using 1,750 fathoms of 3 by 3 rope, 6 fathoms of 9/16-inch ground chain, and four 4-hook sets of half-scale Giffords (with a boat grapnel at the tail end for good measure); see Fig. 1. The average towing speed was 0.85 knot, producing an average dynamometer tension reading of 27 cwt (hundredweight = 112 pounds). The cable was hooked about mid-way along the Giffords, the indication being a rise

Fig. 1. Half-scale Gifford grapnel

Paper 59-153, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 31, 1958; made available for printing December 9, 1958.

C. S. LAWTON is with the Western Union Telegraph Company, New York, N. Y.

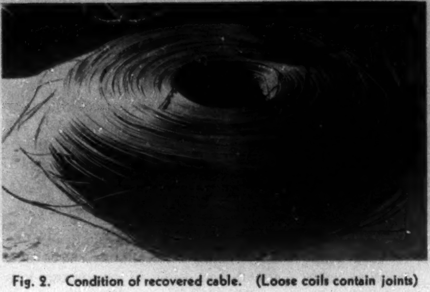


Fig. 2. Condition of recovered cable. (Loose coils contain joints)

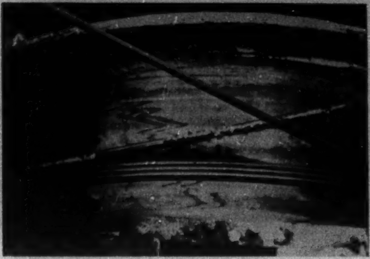


Fig. 3. Parcelling and serving adjacent to joint

of dynamometer tension to 32 cwt. The average speed of raising the grapnel to the surface was 0.31 knot. The tension while doing so dropped from 30 to about 11 cwt. The standing "strain" with both parts of the cable at the bow was 8 cwt.

The short end was found wrapped around the cable for perhaps 50 fathoms when brought to the bow. To be sure of clearing all turns a total of 80 fathoms was picked up on the bight before applying stoppers, cutting off the short end at that point and both parts at the grapnel, thus making it possible to clear the turns and recover the short end separately. This condition was the natural result of the final bight splice dropped at the conclusion of the 1951 laying operation where the nonarmored cable had been spliced directly to the conventional armored cable at the eastern end of the pay-out. (The transitional types had been at the west end.) Upon release of tension, the torsional equilibrium of the armored cable would have been restored by normal twisting, thereby transmitting turns to the adjacent nonarmored cable.

Tension during the actual pick-up which ensued ran between 7 and 9 cwt generally, but with occasional surges to a maximum of about 12 cwt. The picking-up speed averaged about 0.6 knot and while picking up, the downward trend of the cable was aft by 10 to 30 degrees from the vertical. The best condition was found to be about 15 degrees aft, and to our knowledge this is the first time there has been definite substantiation of the theory[8,9] that under certain conditions less tension will be developed if the ship overruns the cable, thereby causing the cable to be inclined forward as it leaves the water. This practice, of course, can be followed only with cable not susceptible to kinking on the bottom when relieved of tension.

Condition of Cable

The main body of the cable came up without any evidence of twisting and coiled down perfectly; see Fig. 2. The polyethylene insulation, which deliberately had been left exposed to worms, if any, showed no evidence of attack except closely adjacent to some of the molded joints where the ship had applied canvas parcelling and untarred hemp spunyarn to elongate the taper in diameter over several feet before laying. These materials as recovered were in poor condition, some frayed or missing entirely, but mostly having to be stripped off for examination of the insulation underneath; see Fig. 3. The evidence was in the form of light scorings, none more than about a thirty-second of an inch in depth but typically at random such as would be caused by teredo attack; see Fig. 4. No worms were found and whether the attack still was in process at the time of recovery is doubtful. It has been suggested that a worm needs support from behind to start the boring action. This would seem to be confirmed. It further has been suggested, in view of the depth of water and proximity of the cable to the Gulf Stream, that these may have been sporadic attacks by stray worms transported by the current and not indigenous to the region at all. Apart from this, the outside of the cable after nearly 7 years submersion could not be distinguished from its condition when manufactured.

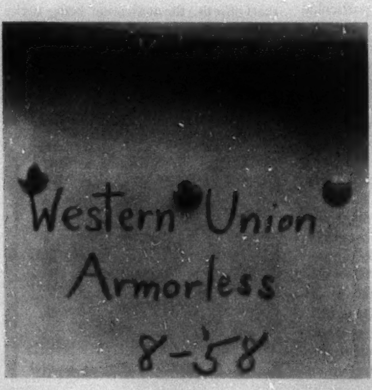


Fig. 4 (left). Nonarmored cable joint—teredo marks accentuated by chalk

Photo courtesy of Bell Telephone Laboratories

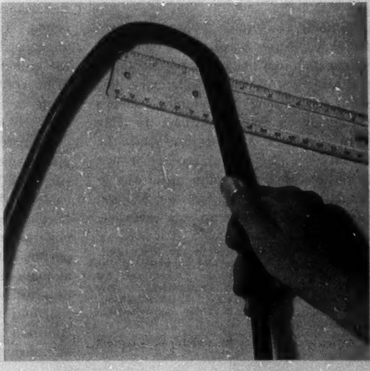


Fig. 5 (right). Cable where hooked and raised by grapnel

One interesting feature of the recovery operation was the bringing of the cable to the surface and inboard over the ship's bow from a depth of over 1,200 fathoms on a Gifford hook, the palm of which was only $2^1/_2$ inches wide, without putting a fault in the cable; see Fig. 5. As a matter of fact, the whole length showed no electrical weakness.

Conclusions

It is evident that a modulus higher than 5 will be necessary for successful recovery in greater depths and that to attain this some use of steel in the strength member is indicated. (When the experimental cable was being designed there was some apprehension about the use of two dissimilar metals since diffusion of water through the dielectric was known to take place and it was thought possible that galvanic action would develop, but this now is recognized as being unlikely because salts do not come through.)

With the successful conclusion of this experiment much of the uncertainty surrounding the ability of nonarmored cable to perform satisfactorily in deep-sea environment has been removed. Naturally, if time was of no importance it would be desirable to make tests in more than one locality and at various depths. Certainly where bad bottom is known to exist with current sufficient to remove the normal deposit of silt, prudence still dictates caution, but it is felt that there now is substantial evidence to support the underlying premise that expensive armor wire on the outside of cable for use in deep water, generally speaking, is unnecessary.

References

1. A Nonarmored Submarine Telegraph Cable, C. S. Lawton, L. H. Hutchins, Jr. *AIEE Transactions*, vol. 72, pt. 1 (*Communication and Electronics*), May 1953, pp. 153–58.

2. Nonmetallic Armored Electrical Submarine Cable, L. H. Hutchins, Jr. *U. S. Patent no. 2,576,227*, filed Dec. 10, 1949, issued Nov. 27, 1951.

3. Improvements in or Relating to Submarine Cable, Eric Willoughby Rollinson. *British Patent no. 667,159*, filed Jan. 12, 1950, issued Feb. 27, 1952.

4. A New Deep-Sea Coaxial Cable, R. A. Brockbank, A. L. Meyers. *Journal*, Institution of Electrical Engineers, London, England, vol. 2 (New Series), 1956, pp. 572–73.

5. Improvements in or Relating to Submarine Cables, R. A. Brockbank. *British Patent no. 703,782*, filed Aug. 16, 1952, issued Feb. 10, 1954.

6. Improvements in and Relating to Submarine Cables, E. W. Smith. *British Patent no. 740,647*, filed Sept. 17, 1954, issued Nov. 16, 1955.

7. A Proposed New Telephone Cable Between the United Kingdom and Canada, R. J. Halsey. *Post Office Electrical Engineers' Journal*, London, England, vol. 50, pt. 2, July 1957.

8. Dynamics and Kinematics of the Laying and Recovery of Submarine Cable, E. E. Zajac. *Bell System Technical Journal*, New York, N. Y., vol. 36, no. 5, Sept. 1957, pp. 1129–1207.

9. A Theoretical Investigation of the Form Assumed by a Submarine Cable During Laying or Recovery, V. G. Welsby. *Monograph no. 274R*, Institution of Electrical Engineers, Dec. 1957

Reperforator-Teletypewriter TT-195( )/FG

D. F. FRICK
NONMEMBER AIEE

Synopsis: Efficient and extremely rapid mechanical energy transfer by means of a design philosophy based on elastic impact of metallic parts made possible the fulfillment of the end function required of a high-speed teletypewriter. Solid-state electronic components enabling data translations and control functions to be performed economically and reliably within minimum space and power limits has proved the final key to practic equipments whose performance far exceeds that of the current standards. The device described in this paper demonstrates the magnitude with which performance might be extended by the marriage of solid-state data-handling techniques with unique mechanical design.

THE ATTACK on the problem of developing a high-speed teletypewriter equipment consisted of extending the electrical signaling and processing of received data as far into the device as practical. The necessity for a mechanical printed and punched paper-tape record obviously requires a mechanical end function, but many decoding and selection operations could be performed electronically, thereby eliminating a large part of the difficulty in achieving high operational speed. A bonus of enhanced reliability, ease of adjustment and maintenance, and greater tolerance to environment and signal distortions were also realized.

The reperforator can be separated into four components providing the following major functions:

1. Selector
   a. Provides signal line isolation from machine ground.
   b. Converts serial 7.0-unit, or longer, Baudot signals, received at one of six discrete transmission rates, into parallel form.
2. Storage Unit
   a. Provides capacity for storing up to 20 5-bit characters.
   b. Address registers and circuitry for insertion of data at constant rates from selector.
   c. Electronic hardware for removal of characters in the order received and delivery upon request by controller.
3. Controller
   a. Converts 5-channel parallel information from storage unit into form acceptable to mechanical unit.
   b. Provides the several electric fundamental outputs to accomplish functions such as ringing signal bell upon reception of sequence FIGS...S, and circuitry to decode shift functions.
4. Mechanical Unit
   a. Together with controller, translates from 5-channel binary information to fundamental (32) outputs.
   b. Punches paper tape with received information.
   c. Provides functions of tape transport, mechanical printing, and upper-lower case shift of typewheel.

The mechanical end function was provided most satisfactorily by utilization of a series of machine elements whose energy and motion were derived by near elastic impact with other machine elements. The usual problems of rebound and vibration were circumvented by using these physical phenomena to advantage within the machine. Figs. 1 and 2 show the reperforator complete and without dust cover respectively.

Controller and Mechanical Unit

A 5-digit binary character is received electrically at the controller. An electromechanical decoding operation is performed resulting in the selection of one of 32 sliding rods. A selected rod presents rigid interference elements in the grooves of a rotating wheel, transferring the selection by means of elastic impact

---

Paper 59-152, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 5, 1958; made available for printing December 8, 1958.

D. F. Frick is with Smith-Corona Marchant Inc., Highland Park, Ill.

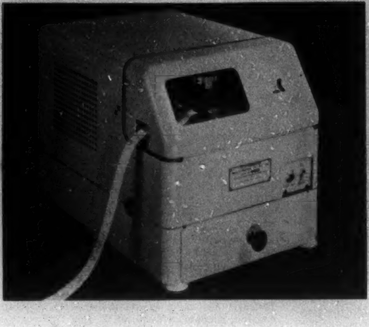


Fig. 1 (left). Reperforator-teletypewriter TT-195( )/FG

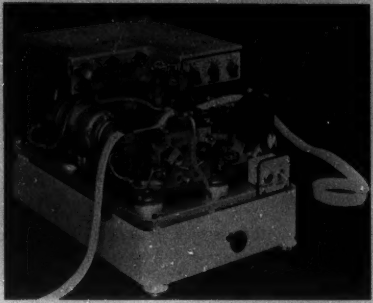


Fig. 2 (right). Reperforator with dust cover removed

to intermediate levers and thence to spring-loaded punches and print hammer. The energy input to the intermediate levers is a function of peripheral wheel velocity, angle of impact between rod and lever, and moment of inertia of the intermediate lever. The latter is adjusted to provide the energy required to perform the desired function, i.e., punch, print, and operate the feed mechanism.

Figs. 3 and 4 illustrate the four locations at which it is possible to select a code rod, each location having eight code rods available, equally spaced about the circumference of a rotating code wheel. These eight shaft positions are monitored by the controller and compared with information from three of the input data bits. The location at which the code rod is to be selected is determined by the states of the remaining two digits of the received data. Each of the four groups of eight rods is displaced 11 degrees 15 minutes of phase from the others. Therefore, the system might be considered equivalent to one wheel containing 32 rods.

The 32-character typewheel is unity geared to the code wheels, resulting in the passage of individual rods past an index point fixed to the machine in registry with each character on the typewheel passing a like index point. A print hammer will be struck by any selected code rod, thereby driving the paper tape into the ink ribbon, into the type face passing the printing position at that time. In addition, each code rod has the five bit code machined into it, interfering with punch operating levers should the rod be selected; see Fig. 5.

The mechanism illustrated in Fig. 6 serves to select the code rods and perform the electromechanical transition at the mechanical unit. When the shaft position monitor has indicated that the desired code rod is about to pass a selection location, an electromagnet removes a latch restraining the "pluck-out" plunger. A leaf spring drives the plunger axially into the path of the passing code rod and the impact resulting from the interference of the ball and cone surfaces of plunger and rod, respectively, drives them apart. The masses of the plunger and code rod are adjusted to provide equal rebound velocities. The impact surface of the plunger consists of a steel ball whose rotation within its socket presents new surfaces to the code rod after each impact; The code rods, likewise, are allowed to rotate freely within the impact wheel. thus, the wear on both ball and cone impact surfaces is distributed over a large area.

The angle of impact of the two elements is a constant independent of the amplitude of interference because of the geometrical shapes of the impact surfaces. The code rod slides axially within the

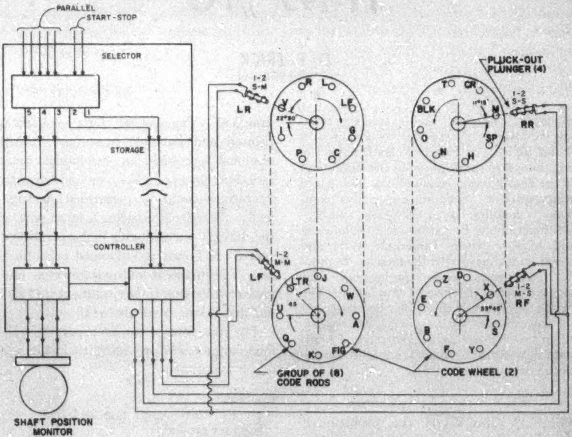


Fig. 3. Electromechanical function

	1–2								
5		●	○	●	○	●	○	●	○
4		●	○	○	●	●	○	○	●
3		●	●	●	●	○	○	○	○
LF	● ●	LTR	U	Q	K	FIG	A	W	J
RF	● ○	X	S	Y	F	B	E	Z	D
LR	○ ●	V	I	P	C	G	LF	L	R
RR	○ ○	M	SP	H	N	O	BLK	T	CR

● Mark.
○ Space.

Vol.
Pt. 1
1959

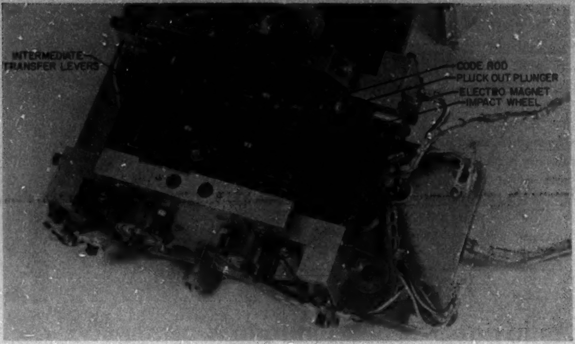

INTERMEDIATE TRANSFER LEVERS
CODE ROD
PLUCK OUT PLUNGER
ELECTRO MAGNET
IMPACT WHEEL

Fig. 4. View of code wheels and rods

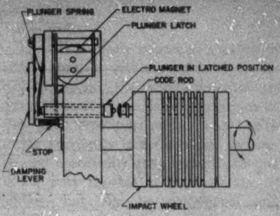

PLUNGER SPRING
ELECTRO MAGNET
PLUNGER LATCH
PLUNGER IN LATCHED POSITION
CODE ROD
STOP
DAMPING LEVER
IMPACT WHEEL

Fig. 6. Code rod selection mechanism

rotating impact wheel 0.050 inch where it is locked by a detent mechanism. The excess energy of the returning plunger is absorbed by impact with the damping lever shown in Fig. 6. The movement of the damping lever is adjusted so that nearly all of the energy of the plunger is transferred to it causing the plunger to halt, and slowly return to rest upon its latch.

The economy of the system now becomes apparent. The high degree of constancy of the energy imparted to the print hammer results in accurate registration of printing from a typewheel rotating at 3,800 rpm. Further, as mechanical energy is derived from elastic impacts (transfer of selection from code rod to punches and print hammer), the printing and punching operations require a minimum of time enabling operational high speed.

The printing energy is derived from a small mass floating freely in the extreme end of the print hammer; see Fig. 7. When struck by a selected code rod, this print hammer travels upward into a solid stop without contacting the recording medium. An aluminum plunger weighing 0.12 gram is allowed to overtravel the print hammer and drive into the paper tape. In this manner, the contact time of the recording medium with the moving typewheel is kept to a minimum, reducing any tendency to smear.

It should be noted that to search out one discrete position of a rotating shaft may require up to one revolution: 16 milliseconds in this case. As the equipment is required to operate at a rate of 75 operations per second (13.3-millisecond period), the reperforator makes use of the fact that some sequences of characters received might provide several selection possibilities in one revolution.

The mechanical functioning requires a three-eighths revolution to complete eliminating the availability of the first two of the eight shaft positions. Therefore, selections every three-eighths revolution are possible should the received sequence of characters be in correct order. Adverse sequences of characters delivered to the reperforator might require $1^1/_4$ revolutions of search. These rotational displacements are equivalent to slightly less than 6 and 20 milliseconds, respectively, whose average is adequately less than the 13.3-millisecond period between received characters. A mathematical determination of the theoretical frequency distribution of these occurrences with random text indicated that were capacity for storing 20 characters provided, prob-

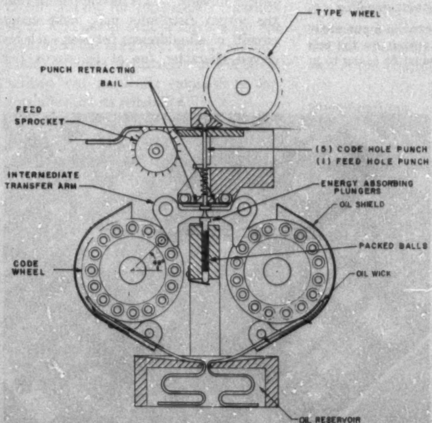

TYPE WHEEL
PUNCH RETRACTING BAIL
FEED SPROCKET
(5) CODE HOLE PUNCH
(1) FEED HOLE PUNCH
INTERMEDIATE TRANSFER ARM
ENERGY ABSORBING PLUNGERS
OIL SHIELD
PACKED BALLS
CODE WHEEL
OIL WICK
OIL RESERVOIR

Fig. 5. Energy transfer and punching mechanism

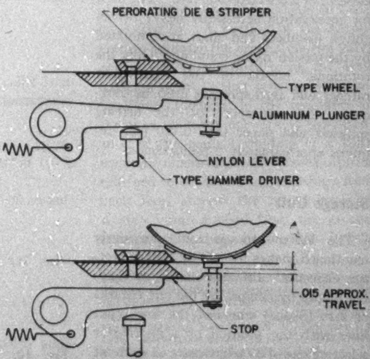

PERORATING DIE & STRIPPER
TYPE WHEEL
ALUMINUM PLUNGER
NYLON LEVER
TYPE HAMMER DRIVER
.015 APPROX. TRAVEL
STOP

Fig. 7. Functioning of print hammer

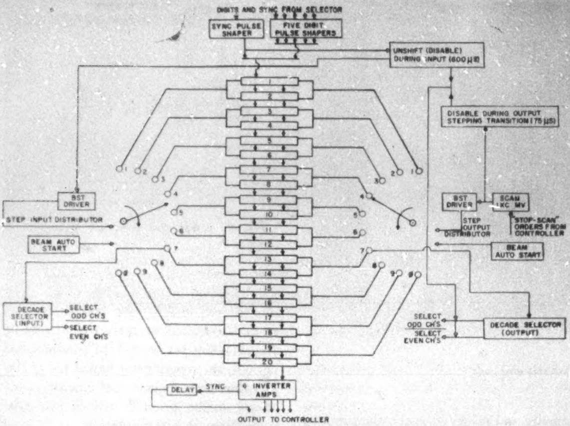

DIGITS AND SYNC FROM SELECTOR
SYNC PULSE SHAPER
FIVE DIGIT PULSE SHAPERS
UNSHIFT (DISABLE) DURING INPUT (600 μS)
DISABLE DURING OUTPUT STEPPING TRANSITION (75 μS)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
BST DRIVER
STEP INPUT DISTRIBUTOR
BEAM AUTO START
DECADE SELECTOR (INPUT)
SELECT ODD CH'S
SELECT EVEN CH'S
BST DRIVER
SCAN 1KC MV
STEP OUTPUT DISTRIBUTOR
"STOP-SCAN" ORDERS FROM CONTROLLER
BEAM AUTO START
SELECT ODD CH'S
SELECT EVEN CH'S
DECADE SELECTOR (OUTPUT)
DELAY
SYNC
6 INVERTER AMPS
OUTPUT TO CONTROLLER

Fig. 8. Storage unit logic diagram

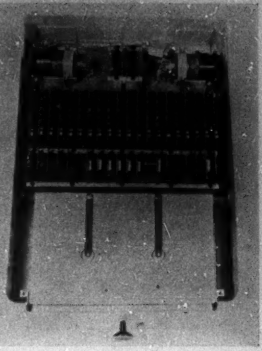


Fig. 9. View of storage unit and chad drawer

ability of overflow in such a buffer would be to the order of one in 10^{-6} operations. The reperforator therefore makes use of an internal buffer that delivers characters to the controller upon completion of each preceding selection at the mechanical unit.

Figures-letters shift functions are provided by driving the 2-row typewheel laterally into these positions upon receipt of the shift functions.

Tape transport is accomplished by a ratchet and pawl mechanism shafted to a sprocket engaging feed holes in the paper tape. Torque is applied to the feed ratchet through a torsion-wound straight-wire spring driven by a felt friction clutch. The restraining pawl is removed from the ratchet by impact with an interference element of a selected code rod. The need for accelerating necessarily heavy driven clutch members and shaft is eliminated by the torsion spring coupling. Only the rotational inertia of the lightweight ratchet and feed sprocket need be overcome. The transport of the driving sprocket and paper tape 0.100 inch requires approximately 2 milliseconds.

Storage Unit

The buffer storage unit previously mentioned makes use of 120 binary storage elements. Information is read into the device by comparing the one (Mark) or zero (Space) states of the five input lines with the position of a distributor selecting one of 20 addresses. A sixth bit is stored at each address to facilitate logical removal of characters. The 20 6-bit addresses are available to the 10-position distributors by arranging the character addresses in two decades selected by a single binary stage triggered on each revolution of its associated distributor. The input distributor progresses forward, one address at a time, upon receipt of characters from the reperforator's selector. The output distributor, however, has more stringent requirements to cope with:

1. Characters must be removed in the exact sequence received.
2. It should not be necessary to switch out the storage unit when operating at low speeds or during intermittent signaling.
3. Since the relation between input occurrence and demand for output by the controller is random, care must be taken to insure that characters are not lost or repeated when in-out functions occur simultaneously.

In short, a device is required whose function is that of a "rubber register"—a storage medium that has no capacity when empty, and increases its capacity with filling.

The output distributor is therefore instructed to scan the entire storage unit for the presence of characters while the controller has no data to digest. If a character is inserted in the storage unit, the arrival of the output distributor at that address will cause the controller to act upon that information, ordering the distributor to cease its search. Each time a selection is completed within the controller, the storage unit obeys the order to scan forward. It is then apparent that the output distributor may make many circuits of all addresses between each received character (as at 60 words per

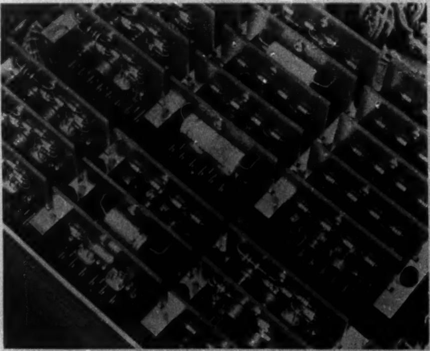


Fig. 10. Modular components of controller

minute), or it may merely "search" forward one address at a time during those periods when several characters remain in storage. Fig. 8 illustrates, in block form, the functioning of the storage unit. Fig. 9 shows the electronic hardware comprising this component.

All logical functioning is accomplished by means of resistor-diode gates and low-level switching transistors. The electromechanical transition at each of the four code rod selection locations is provided by high-speed electromagnets driven by power transistors. A total of 183 germanium transistors and 660 diodes are employed in the reperforator. The unit requires 175 watts of a-c power, weighs 55 pounds, and occupies 1.5 cubic feet of space. The electronic section is constructed of plug-in modular components, each contributing one of several standard logic functions. Fig. 10 shows a group of such modules comprising the controller portion of the reperforator.

Conclusions

The use of elastic impact as a means of energy transfer within a mechanical system has opened the way to practical accomplishment of mechanical functions at very high repetition rates. The equipment previously described has demonstrated specifically that impact printing is not only feasible at high operational speeds, but can be economical and reliable as well.

A Magnetic Amplifier Flow Controller

H. E. DARLING
MEMBER AIEE

Synopsis: An all-magnetic amplifier controller is described. The design has the long life and dependability expected from an all-magnetic circuit and still meets the exacting accuracy and stability requirements of industrial process control. It is intended primarily for fast response (flow) processes, and is designed for 10–50-ma d-c input and output.

AS THE PRESSURE for more efficient production and more uniform products in industrial processes increases, so does the need for more sensitive and reliable instrumentation designs. To meet this demand, industrial instrument manufacturers have been developing electronic control systems, and one such system uses 10–50-ma (milliampere) direct current for intelligence transmission. One component of this system, the controller, has been built which uses magnetic amplifiers exclusively. It will be described with particular emphasis on factors affecting its reliability.

In industrial process control, a large percentage of the controllers are analog computers requiring time functions for integration or differentiation. Since this is normally accomplished by d-c circuitry, the adoption of d-c information transmission made it seem logical to use an all d-c amplifier as a major component in the controller. Magnetic amplifiers which are basically d-c amplifiers, have been developed in recent years to a point where they are the most reliable of all methods of electronic amplification, and within their limits of application, should be ideal for this type of industrial controller design.

The controller to be described is intended primarily for fast response systems such as those encountered in flow control problems. The Foxboro Company has examined the problem of liquid flow in detail[1–4] which led to the development of the well-known *M/59* pneumatic flow controller. Extrapolating from pneumatic experience, a 2-term electronic controller having a proportional band (pb) adjustable from 100–500%, reset times adjustable to 30 seconds per repeat, and no derivative, would be expected to perform in an electronic control system in a manner strictly analogous to the *M/59* in its pneumatic system. From the characteristics of the *M/59* were obtained the original specifications for the all-magnetic amplifier controller.

Factors affecting reliability in the design of magnetic amplifiers for industrial processes has been previously reported,[5] and have been applied throughout the design of this controller.

Specifications

The magnetic amplifier controller is to have the following general characteristics.

1. Proportional band: 100–500%; (10–50-ma input versus 10–50-ma output at 100% pb.)
2. Reset time: variable from 1–30 seconds per repeat.
3. Rate: none.
4. Power supply: 60 cycles, 117 volts, regulated ±2%.
5. Linearity: 0.5%.
6. Zero stability: 0.5%.
7. Adjustable set point: 0–100% scale, accurate ±0.25%.
8. Zero frequency gain: 200 at 100% pb.
9. Frequency response: down 4 db (decibels) at 10 cps (cycles per second).
10. Ripple voltage in output: 1% of full scale voltage.
11. Provided with both manual and automatic operation controls, and arranged for bumpless transfer.
12. Provided with set-point error indicator, and valve position indicator.

Principles of Operation

The block diagram of the controller as used for automatic operation is outlined in Fig. 1. Direct current from a process transmitter is applied to the input terminals of the controller, and the voltage developed at these terminals is opposed by a reference voltage obtained from a precision Zener diode source, the latter serving as a set point. The difference b tween the control and set-point voltage attenuated by the pb adjustment an applied to the input stage of the m amplifier *SSR-1*.

The main amplifier consists of tw caded stages of self-saturating mag amplification *SSR-1* and *SSR-2* w a d-c output, and with the ripple co at of this output reduced by a filter to a low level (0.5%). A portion of the output voltage actuates the process control valve and a portion is applied to an *R-C* time constant network that provides the reset function. Included in the resistance of the time constant network is the resistance of the control winding of a third magnetic amplifier *SSR-3* whose output is returned to the first stage, creating a feedback loop around the main amplifier. A step input applied from the process transmitter is then characterized by the

---

Paper 59-157, recommended by the AIEE Recording and Controlling Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted July 21, 1958; made available for printing December 23, 1958.

H. E. Darling is with the Foxboro Company, Foxboro, Mass.

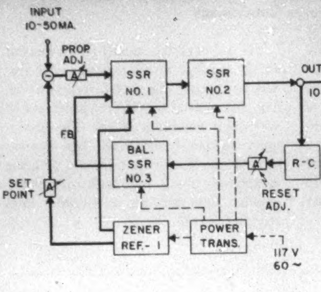

INPUT
10-50MA.
PROP. ADJ.
A
SSR NO. 1
SSR NO. 2
OUTPUT
10-50 MA.
FB
BAL. SSR NO. 3
A
R-C
RESET ADJ.
SET POINT
A
ZENER REF.- 1
POWER TRANS.
117 V 60 ~

Fig. 1 (left). Magnetic amplifier controller, automatic operation

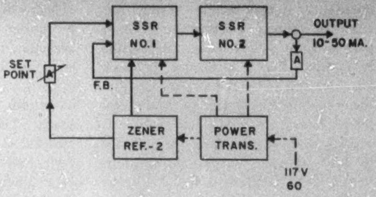

SET POINT
A
SSR NO. 1
SSR NO. 2
OUTPUT
10-50 MA.
A
F.B.
ZENER REF.- 2
POWER TRANS.
117 V
60

Fig. 2 (right). Magnetic amplifier controller, manual operation

feedback system to produce a proportional step output followed by reset action.

The main amplifier may also be used to operate the process by manual control, as shown in the block diagram of Fig. 2. The reset circuit and the transmitter input are by-passed through a conventional resistance feedback network, and this feedback voltage is opposed by a second Zener diode reference source, which constitutes a manual positioning adjustment for the unit. In this condition of operation, the amplifier is a stabilized system of 100% pb, and its output valve current follows accurately the settings of the manual set-point voltage.

This method of manual control is only one of several possibilities. For instance, instead of using a common amplifier, the manual and automatic systems could be separate units which would facilitate interchangeability and service of units in the field. Such an approach is under consideration.

Switching means are provided for changing from automatic to manual control and vice versa without upsetting the process, as will be described later.

Description of Amplifier Details

A complete schematic diagram of the controller will be found in Fig. 3, and a description follows.

The input stage of the main amplifier SSR-1 is of the parallel connected self-saturating type, and is provided with two identical control windings to allow electrical isolation between input and feedback circuits. Two additional control windings are required for bias and antihunt purposes.

A center-tapped transformer winding energizes the gate circuits of this stage, requiring the use of only two diodes, thereby reducing to a minimum the number of components used in the gate circuit. Silicon-fused junction diodes are used because of their remarkable electrical and mechanical properties, and their long life expectancy. The load of this gate circuit is the control winding of the second stage.

The 10–50-ma direct current from a process transmitter, when applied to the input terminals of the controller, flows through a 100-ohm precision wire-wound resistor, generating an input voltage be-

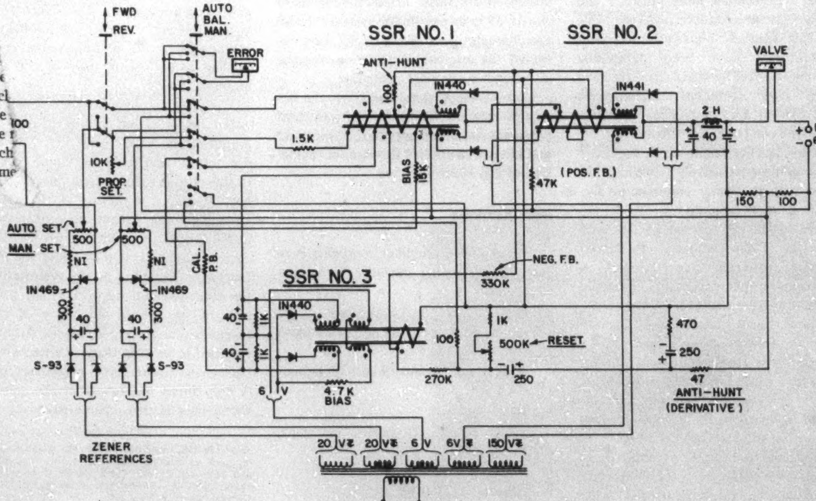

FWD
REV.
AUTO
BAL.
MAN.
ERROR
SSR NO. 1
SSR NO. 2
ANTI-HUNT
100
IN440
IN441
VALVE
2 H
40
+ LOAD
- 600Ω
100
1.5K
10K
PROP. SET.
BIAS 15K
47K
(POS. F.B.)
150
100
AUTO. SET
MAN. SET
500
500
NI
NI
IN469
IN469
300
300
40
40
CAL. P.B.
NEG. F.B.
330K
SSR NO. 3
IN440
40
40
1K
1K
100
1K
500K
RESET
470
250
270K
250
47
ANTI-HUNT (DERIVATIVE)
S-93
S-93
6 V
4.7K BIAS
ZENER REFERENCES
20 V
20 V
6 V
6V
150 V
120V -60~
REGULATED ±2%

Fig. 3. Magnetic amplifier controller

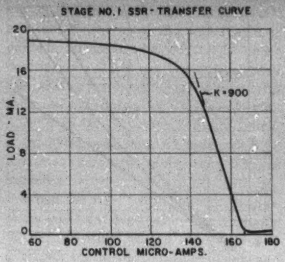


Fig. 4. Stage no. 1 SSR, transfer curve

tween the limits of 1–5 volts. This input is opposed by a voltage derived from a Zener diode reference source, whose output is manually adjustable by means of a potentiometer from 1–5 volts, and the difference between the two voltages is applied to one of the control windings of the input stage in series with the pb potentiometer. This variable reference voltage is the controller automatic set point, and the difference between the set-point voltage and the process voltage is indicated by a sensitive microammeter when the controller switch is in the automatic operation position. Its purpose is to indicate the deviation of the process from the desired control point.

Bias for this first stage, which must be regulated to a high degree of precision, is derived from a Zener source. The bias current is adjusted to the required value by a series resistor which also acts to prevent the circulation of gate circuit induced voltages in this bias winding. The magnitude of the limiting resistor is many times larger than the ohmic resistance of the bias winding, and because the resistor is of the type that has a small temperature coefficient, the effect of temperature of the bias current is minimized.

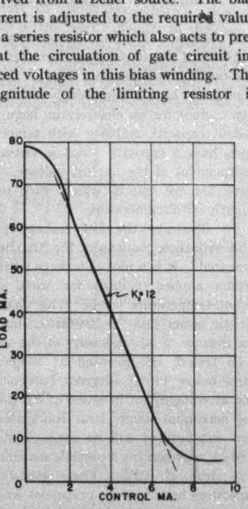


Fig. 5. Stage no. 2 SSR, transfer curve

The transfer curve for this stage (Fig. 4) shows a current gain of 900 in its linear operating region. Further construction details are described in the appendix.

The second stage of the main amplifier SSR-2 is also of the parallel-connected self-saturating type, and is provided with a control, antihunt, and decoupling winding. Maximum reliability again dictates that the gate circuits be supplied from another center-tapped transformer winding, using only two diodes of the silicon-fused junction type.

The output of *SSR-2*, being a chopped-sine waveform, requires a high degree of filtering, and is accomplished by an *L-C*π type filter which reduces the ripple to less than 1%. High-reliability electrolytics were chosen for filter condensers. The output current flows through an indicating ammeter to show valve position, the external load of a valve and other accessories (600 ohms), and then through two precision wire-wound resistors of 100 and 150 ohms respectively. These latter resistors are used for feedback purposes.

With zero control signal into *SSR-1*, its bias is used to adjust the interstage circulating current between it and *SSR-2*, and this current provides the necessary bias for the output stage. By this means, a bias source for *SSR-2* is eliminated, and adds to the reliability of the system.

The transfer curve of *SSR-2* (Fig. 5) shows a current gain of 12. Further construction details are described in the appendix.

With the two stages connected in cascade, the over-all open-loop transfer curve for the amplifier is shown in Fig. 6. An input current change of 3.5 μa (microampere) (3×10^{-9} watts) is required for a 10–50 ma output current change, or the system has an open-loop current gain of 11,000 and a power gain of 87 db. The open-loop frequency response of the main amplifier and its output filter is shown in Fig. 7, indicating that the amplifier and its associated filter has a long time constant.

When used in the automatic control condition, the voltage drop across the 150-ohm resistor in series with the output is applied to an *R-C* time constant network. The purpose of this *R-C* network is to provide the required resetting action of the controller, and also to introduce a fixed amount of derivative for over-all amplifier stabilization.

Specifications of the controller require reset times of 30 seconds per repeat, which is equivalent to saying that the time constant of the reset circuit should be 30 seconds. In vacuum-tube circuitry, the voltage developed across a resistor in series with a reset condenser would be applied to the grid of the tube as a feedback voltage since the vacuum tube constitutes essentially no load upon this resistor, and the feedback circuit then becomes quite simple. In the case of magnetic amplifier, the condition is different because it is difficult to build a magnetic amplifier of sufficiently high input impedance so that serious loading on the reset resistor would not occur. However, it is relatively easy to build a magnetic amplifier that is capable of measur-

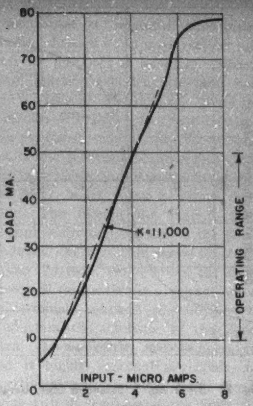


Fig. 6. Main amplifier open-loop transfer curve

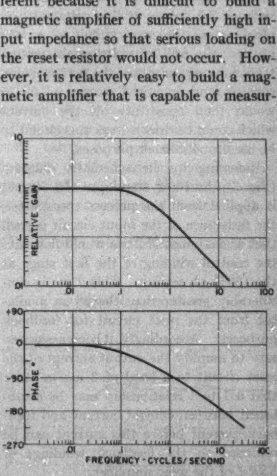


Fig. 7. Main amplifier open-loop frequency response

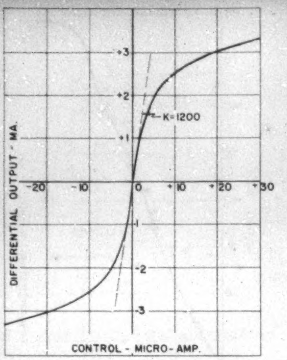


Fig. 8. Stage no. 3 balanced SSR, transfer curve

ing a very small current, and by measuring current flow through the reset resistor rather than the voltage across it, a method is found which is suitable for a magnetic amplifier controller.

In order to get a condenser current flow in the reset circuit of a suitable value, it becomes necessary to use a large capacitor and relatively small resistance in the *R-C* circuit. For the controller being described, a 250-μf (microfarad) capacitor and a reset resistor of 0.27 megohms was chosen. At full-scale output (50 ma), the maximum voltage which would be developed across a 150-ohm feedback resistor is 7.5 volts. If this voltage is applied across the reset circuit as a step change, the full 7.5 volts would instantaneously appear across the 0.27-megohm resistor, and an initial current of 28 μa would then constitute all the current which could be drawn from this circuit to be used for feedback purposes.

Referring to the schematic diagram (Fig. 3), if a 100% step input (10–50 ma) is applied from the process transmitter, the resistance of the input circuit is such that a maximum of 2 ma would flow into the control winding of the first stage at the 100% pb setting. This 2 ma is considerably greater than the 28 μa available from the reset circuit for feedback purposes. Accordingly, it becomes necessary to amplify the output current of the reset circuit to a level of 2 ma in order that a 1 to 1 relationship may be established between control current and feedback current before this current can be returned to the feedback windings of *SSR-1* of the main amplifier.

Adjustment of reset time constant is accomplished by shunting the 0.27-megohm resistor and input winding of *SSR-3* with an adjustable potentiometer. This by-passes a portion of the reset capacitor current and decreases the reset time. This method was chosen because changes in the potentiometer resistance have no effect on the gain of the reset current amplifier, which would change the proportional band of the system. The reset and pb adjustments are virtually independent. At its maximum setting, this potentiometer reduces the maximum current flowing into the control winding of *SSR-3* from 28 to 18 μa.

The amplifier to be used to step up the 18 μa available to the 2-ma level necessary for feedback, must satisfy several stringent requirements. It must be able to sense current of the order of 0.25% of 18 μa, or 0.04 μa (1.1×10^{-12} watts). It must have a zero stability with time and temperature which must not exceed the equivalent input signal change of 0.04 μa, its noise level must be below this same figure, and it must have a stable current gain of 110. It must also be of the balanced type; i.e., for zero input current, the output current must also be zero, and the output must reverse in polarity in accordance with a polarity change of the input signal. If the zero of the amplifier should drift, this would appear to the main amplifier as though the process or the set point had been altered, whereas a change in gain of the amplifier will effect the pb of the system only. Because of the low level at which this reset current amplifier operates, all of the details of its construction become critical.

The special amplifier devised for this application consists of two half-wave self-saturating magnetic amplifiers in push-push combination (Fig. 3). Both gate circuits fire simultaneously on the same half of the 60-cycle signal, and are reset on the opposite half by the bias winding. Each gate circuit supplies current to its own individual load resistance, and the output of the amplifier is taken differentially across these resistors. The amplifier is of the balanced type whose output polarity is directly related to the polarity of the control signal.

The transfer characteristics of this amplifier *SSR-3* working into a load resistor equal to that of one of the 1,000-turn windings of the first stage of the main amplifier, is shown in Fig. 8. The transfer curve is symmetrical with respect to the origin, has a maximum gain of 1,200, and an effective gain at 2 ma output of 400. This amplifier was deliberately made to have a current gain several times larger than that actually required in order that some of this gain could be used degeneratively to linearize and stabilize it. A more detailed description of the construction of this amplifier will be found in the appendix.

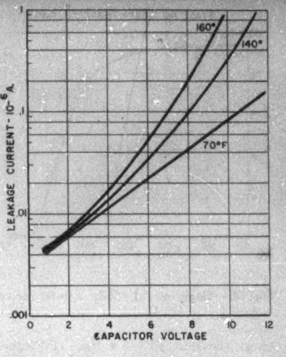


Fig. 9. Leakage current, tantalytic porous plug capacitor

It has been emphasized that this reset current amplifier must have excellent zero stability with time and temperature. It is equally important that the reset condenser used in the time constant circuit has a very small leakage current, because this leakage current when amplified by the reset current amplifier, would be returned to the input of the main amplifier and would appear to that amplifier as though either the control point or process had changed. The reset condenser has a capacity of 250 μf and is an electrolytic of the tantalum porous plug design. These condensers have a surprisingly low leakage current for an electrolytic, have excellent capacity stability with temperature, have a capacity which is virtually independent of the applied voltage, and have leakage effects which vary only slightly with temperature.

To illustrate the low leakage possible with these condensers, Fig. 9 has been prepared. It is a plot of leakage current versus applied voltage for three different temperature levels. This characteristic is not just one condenser, but is the average of 30 condensers of the same type tested. If operated at temperatures below 140 F (degrees Farenheit), and at voltages not in excess of 7.5 volts, the maximum steady state leakage current encountered will be approximately 0.06 μa, representing a possible maximum zero drift of 0.3%. These tantalytic capacitors have leakage properties which approach those of the best paper-foil units available at the present time (having an *R-C* factor of 63,000 megohm-μf), and it is by virtue of this fact that it is possible to make the controller under description. They also have long shelf life, and several years' experience with their use indicates a high degree of reliability. In the controller actually tested, a capacitor was selected from which a maximum zero drift of less than 0.2% is expected.

When the reset loop is closed, the amplifier will hunt due to the time constants of the individual stages used. This is overcome by the use of rate feedback, and is accomplished in two separate paths. The first path is an interconnection between properly polarized windings of *SSR-1* and *SSR-2* of the main amplifier through an isolating resistor (Fig. 3). This method of feedback utilizes circulating second harmonic currents generated by the first stage of the amplifier which flow into the second stage. It is a simple, reliable, and effective way of modifying the time constant of an amplifier without the use of an *R-C* circuit, but can only be used where an even number of self-saturating stages are employed. It also reduces the forward gain of the main amplifier slightly. A small amount of positive feedback around *SSR-2* overcomes the loss of gain. This feedback loop corrects only for the time constant of the two stages of the main amplifier.

A second source of a derivative signal is required for the reset current amplifier, and is obtained by the use of a small amount of resistance-capacity placed ahead of the reset capacitor in the manner customarily used for rate feedback (Fig. 3). Tantalytic capacitors must be used for reasons of stability.[5] With these two feedback circuits properly adjusted, the controller becomes stable under a step upset of any magnitude.

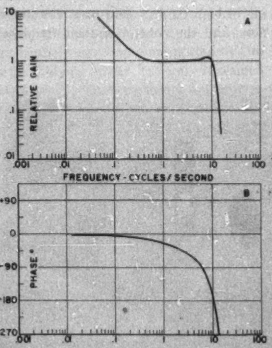


Fig. 10. Frequency response at 100% pb, 1 second repeat

A 3-position switch is provided in the controller in order that bumpless transfer may be made from automatic to manual control or vice versa. In normal automatic operation, the error indicating microammeter is connected differentially between the automatic set-point voltage and the input voltage of the process. When one wishes to transfer to manual control, the switch is placed in its middle position, which leaves the controller on automatic operation but places the error indicator differentially across the manual set-point voltage and the feedback voltage from the 100-ohm resistor in the output circuit. The manual set point is then adjusted until the error indicator reads zero. Under these conditions, the switch can now be turned to its third position, or that of manual control, and since the difference between the feedback voltage and reference voltage is zero, no change in the valve output current will result by this switching operation. The error indicator is at the same time returned to the automatic condition.

Once the switch is placed in the manual position, adjusting the manual set potentiometer will change the valve to any desired position. The amplifier is now a 1 to 1 proportional amplifier with a high degree of degenerative feedback around it and the valve current will faithfully follow the manual set-point adjustment.

When transferring from manual to automatic operation, the automatic set point is adjusted until the error indicator reads zero at which time the switch can be thrown from the manual to the automatic position without disturbing the valve current. By this simple method of switching, bumpless transfer can easily be made from automatic to manual or vice versa.

Reference has been made to the use of Zener systems as a source of constant voltage for both automatic and manual set-point adjustments. These Zener systems use a silicon junction diode having the property that when a voltage is applied across its terminals in the reverse direction, it shows a high resistance until a certain critical voltage is reached, at which point the device rapidly becomes a conductor. Its action is very much like that of a gas regulator tube which is nonconductive until the striking voltage is reached.

These diodes when used in a proper circuit, constitute a precise and reliable source of constant voltage. As shown in Fig. 3, each Zener source obtains its d-c power from a separate transformer winding through a full-wave rectifier and filter for purposes of isolation. Silicon-fused junction rectifiers are used here also for reasons of reliability. Since the controller is designed to operate from a 2% regulated supply, the filtered output of the rectifiers is also regulated to this same value, and in combination with the Zener diodes, forms a 2-stage regulator. The filtered d-c current flows through a limiting resistor to the Zener diode and the fixed load across the Zener diode. At room temperature, the Zener diode will now maintain a voltage across its terminals constant to 0.03%. The conduction characteristic of the Zener diode varies slightly with temperature and is highly reproducible; thus, by inserting a small amount of nickle wire in series with the load across the diode, the temperature effect can be reduced to a very small value. Regulators of this sort have been running for over 3 years with no measurable change in their properties, and have proved to be a precise and reliable source of reference voltage.

Experimental Results

All tests of gain and zero were made using a battery source for input current. Input and output currents were measured with a Rubicon potentiometer as a voltage drop across a precision 10-ohm resistor. In this manner, current could be read in milliamperes directly from the Rubicon scale to four significant figures. A 600-ohm load resistor was used, except as otherwise noted, and the a-c supply to the controller was regulated to ±2%, with an a-c waveform of less than 5% harmonic distortion.

Frequency response measurements were made using a Hewlett-Packard low-frequency oscillator as an input sine-wave source, and both input and output were recorded on a calibrated Brush oscillograph.

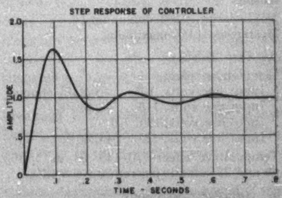


Fig. 11. Step response of controller

Reset time was measured using the

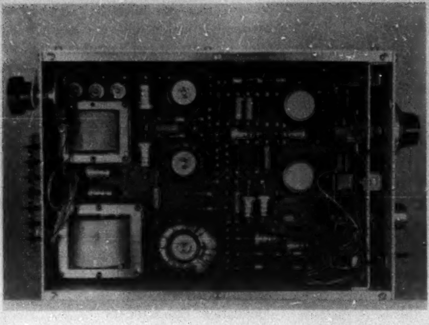


Fig. 12. Research model, magnetic amplifier controller, side view

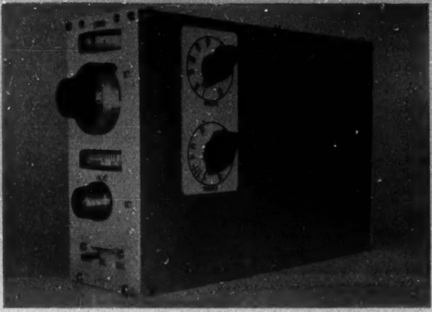


Fig. 13. Research model, magnetic amplifier controller, oblique view

Brush oscillograph. The balance of the controller was upset by a 10% step input from the battery, and the time required for the output to double was read from the chart.

Zero frequency gain was determined by balancing the controller at 0% valve setting using minimum reset time, and measuring the input change required to have the controller rebalance at 100% valve setting.

Ultimate sensitivity was determined at 50% valve setting by observing the smallest change of input that would cause the controller to drift away from the 50% point. This was so small that only an estimate of its value could be made.

Elevated temperature effects were determined by placing the entire controller in an oven at a temperature 70 F above room ambient, and measuring changes in zero, span, and reset times at the elevated temperature.

TEST RESULTS

1. *Proportional Band:* 100–1,000% at 70 F.
Effect of temperature rise (70–140 F) on 100% pb at maximum reset: +0.6%.

2. *Reset Time:* Variable from 1–35 seconds per repeat.

Effect of temperature rise (70–140 F) on reset time at 35 seconds per repeat: 0.

3. *Linearity:* At 100% pb, input versus output.
Nonlinearity, 0.2% maximum
Hysteresis, 0.1% maximum

4. *Zero Stability.*
Line voltage effect.
  Input line volts, 105... −0.25% change
  Input line volts, 115... −0.025%
  Input line volts, 120...0
  Input line volts, 125... +0.05%
Temperature effect (70–140 F) at 100% pb and 35 seconds per repeat: +0.13%.

5. *Adjustable Set Points.*
Effect of temperature (70–140 F) on Zener diode voltage:
  Automatic set, +0.07%
  Manual set, +0.13%
Maximum fluctuation during 24-hour interval at 70 F: 0.03%.

6. *Zero Frequency Gain:* 425 at 100% pb.

7. *Ultimate Sensitivity:* Approximately 0.03% at 100% pb.

8. *Load Sensitivity:* (Change at 100% valve current).
Load, 500 ohms... −0.125%
Load, 600 ohms...0
Load, 900 ohms... +0.15%

9. *Ripple Voltage in Output:* 0.6% peak to peak maximum.

10. *Frequency Response:* (See Fig. 10).
Essentially flat to 6 cps; down 4 db at 10 cps.

11. *Step Response:* At 100% pb and 35 second per repeat; see Fig. 11
Rise time: 0.09 second with 60% overshoot into 600-ohm resistive load
Transport lag: 0.02 second

The controller was tested on the Foxboro liquid flow system[4] using a newly developed electro-hydraulic valve operator and an electronic d/p cell transmitter to form an all-electronic control loop. With a 200% pb and reset of 1.0 second per repeat, operation was found to be virtually identical with that of the analogous *M/59* pneumatic control.

A research model of the controller is shown in Figs. 12 and 13. In these, the general size and type of components used are clearly shown.

Fig. 14 is the production version of the controller package. It is designed for panel mounting, is of the plug-in type, and slides out of its case for adjustment of proportional band and reset. The magnetic amplifier is a plug-in subassembly.

Conclusions

The successful construction of an all-magnetic amplifier controller is the result of the development of a very low power input level magnetic amplifier, the high degree of development of the silicon junction rectifier, and the excellent low leakage properties of tantalum electrolytic capacitors. The successful cascading of high current gain magnetic amplifiers, and the control of their frequency

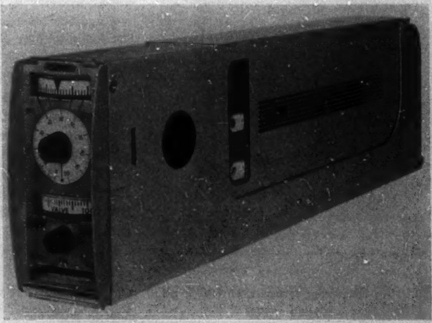


Fig. 14. Production package

response characteristics by feedback techniques makes possible the degree of overall constancy of zero, gain, and time constant required for a precision device. All this has been accomplished without the use of any component of short life expectancy; hence, from the controller described, may be expected a high degree of reliability.

Appendix

The core packages used for both *SSR-1* and *SSR-2* were selected for those properties which yield the ultimate in reliability, and construction details were employed in winding and encapsulating the finished unit with accent on reliability, using the principle reported in reference 5. Winding details are as follows.

SSR-1

Core: 2-mil Hymu *80* tape, matched pair.
  Area = 0.128 cm^2 (centimeter)
  Core length: 6.74 cm
Gate: 1,750 turns no. 34 copper.
Input: 1,000 turns no. 40 copper.
Feedback: 1,000 turns no. 40 copper.
Bias: 500 turns no. 40 copper.
Antihunt: 500 turns no. 40 copper.

SSR-2

Core: 4 mil Orthonol, matched pair.
  Area = 0.544 cm^2
  Core Length: 9.84 cm
Gate: 3,500 turns no. 32 copper.
Input: 250 turns no. 34 copper.
Antihunt: 1,000 turns no. 40 copper.
Decoupling: 12 turns no. 22 copper (short-circuited).

The reason for the use of the short-circuited winding on *SSR-2* has been previously reported.[6,7] These two amplifiers are very similar to those described in these references, although used for a quite different application.

The balanced amplifier *SSR-3* is described in reference 5. Winding details are as follows.

SSR-3

Core: 6 mil Hymu *80* stamped rings, matched pair.
  Area = 0.172 cm^2
  Core length = 5.8 cm
Gate: 1,000 turns no. 34 copper.
Bias: 200 turns no. 34 copper.
Control: 2,500 turns no. 42 copper.

References

1. FACTORS IN PRECISE CONTROL OF LIQUID FLOW, **A. R. Catheron.** *Instruments*, Instrument Society of America, Pittsburgh, Pa., vol. 24, no. 6, June 1951, pp. 705–10.

2. CONTROL VALVE CHARACTERISTICS AND THEIR EFFECTS ON LIQUID FLOW CONTROL, **R. Milham, A. R. Catheron.** *Symposium*, Instrument Society of America, Newark, N. J., Apr. 1952.

3. DEVELOPMENT OF A UNIVERSAL CONTROLLER FOR LIQUID FLOW, **H. L. Bowditch.** *Paper no. 52-15-3*, Instrument Society of America, Mar. 1952.

4. DYNAMICS OF LIQUID FLOW CONTROL, **A. R. Catheron, B. D. Hainsworth.** *Industrial and Engineering Chemistry*, Washington, D. C., vol. 48, no. 6, June 1956, pp. 1042–46.

5. MAGNETIC AMPLIFIERS FOR RELIABLE PROCESS CONTROL, **H. E. Darling.** *Proceedings*, Oak Ridge National Laboratory Conference on Nuclear Process Instrumentation and Control, Gatlinburg, Tenn., May 20–22, 1958.

6. A DIRECT CURRENT TRANSFORMER WITH ADJUSTABLE RATIO, **H. E. Darling.** "Conference on Magnetic Amplifiers," *AIEE Special Publication T-98*, Aug. 1957, pp. 236–44.

7. D-C TRANSFORMER HAS CONTINUOUS ADJUSTABLE RATIO, **H. E. Darling.** *Electronic Design*, New York, N. Y., vol. 6, no. 5, Mar. 1958, p. 40.

Taut-Band Suspensions for 250-Degree Instruments

V. S. THOMANDER
MEMBER AIEE

R. C. MacINDOE
ASSOCIATE MEMBER AIEE

THE taut-band suspension instrument contains a moving element suspended by two metal bands under tension which replace the conventional spiral springs, jewels, and pivot bearings. (The term "taut-band suspension instrument" is used in this paper as reference to an instrument element that is supported by two metal ribbons under tension, sustained by springs.) This new design is free from bearing friction, thus eliminating the major causes of trouble experienced with conventional pivot and jewel bearings such as pivot wear and flattening, or damaged and dirty jewels.

The principal objective that motivated the design and development of the taut-band suspension instrument was the elimination of wear friction between the pivots and jewels. In all conventional switchboard and portable electric indicating instruments employing pivots and jewels, pivot wear and flattening resulting from service and handling has always been a major problem. To keep friction to a minimum: 1. the pivot and jewel cone points have been made with smaller and smaller radii, 2. much has been done to make the pivots of harder and tougher materials, and 3. the surfaces of each are honed and polished to a sheen-like finish. Further improvements have been made by designing moving elements as light in weight as practical. For some instrument designs the jewels have been backed up with a low-pressure spring which permits the jewel to retract under shock, thus allowing a bumper stop to take up the main part of the impact.

Even with the advances in these areas, the remaining friction problem that inevitably exists still places limits on instrument design. In service there still is unavoidable wear which increases friction, with resultant maintenance expense and precautions. Under severe vibration conditions, such as are encountered on railroads and certain types of machines, pivots and jewels have to be replaced periodically.

First Taut-Band Suspension

The obvious solution to the problem of pivot-jewel friction and wear is to try to eliminate it with a frictionless and wearless type of mechanism. An instrument element suspended by strips at high tension was actually used for a short time in Aryton and Perry's D'Arsonval ammeter in 1879.[1,2] It was stated that this type of suspension was "not advisable unless the coil is balanced with exceeding care by the maker, or there is special provision for balancing, for it is difficult to attach the suspension wires so that their axes pass through the center of gravity of the coil when the suspension is drawn taut." There were, of course, other problems and limitations. The elements were too heavy to be supported horizontally by the tautness of the suspension, the suspensions were long, and no provisions were made to prevent breakage from handling or mechanical shock. At best this suspension could be used only in the vertical position.

As the requirement for switchboard instruments expanded into every field of industry, the need for horizontally mounted instruments in many applications became

---

Paper 59-159, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 30, 1958; made available for printing December 12, 1958.

V. S. THOMANDER and R. C. MACINDOE are with the Westinghouse Electric Corporation, Newark, N. J.

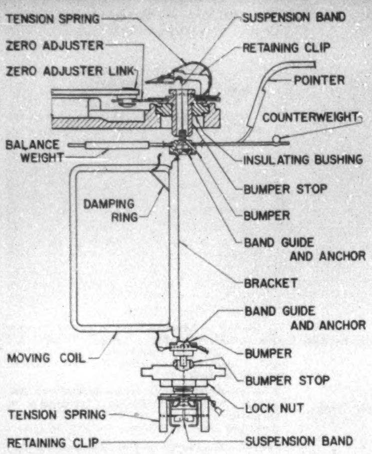


Fig. 1 (left). Taut-band suspension moving element assembly

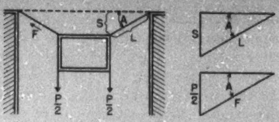


Fig. 2 (right). Sag in a taut-band suspension element in the horizontal position

essential. This may have been a major reason for the concentration on and common use of the pivot-jewel type of mechanisms for instrument moving elements. In any event, pivot-jewel bearings have been standard in instruments for about three-quarters of a century. The continuing research and development on these bearings eventually began to reach the point of diminishing returns, even though the friction problem was not completely solved.

In recent years, European manufacturers[3] have redesigned some of their 90- and 100-degree scale instruments to employ a form of taut-band suspension element. In the United States, the Greibach Instrument Corporation developed a bifilar suspension element. This type of suspension operates on a different principle than the taut-band suspension and therefore will not be discussed in this paper. This paper concerns itself with the taut-band suspension element (Fig. 1) developed by the Westinghouse Instrument Department.

Elements Must Meet Severe Requirements

Recent developments in instrument manufacturing techniques and new materials were major factors in making it possible to develop a taut-band suspension element that would meet the severe standards of performance desired. Other major considerations were that the tension in the band and the element weight be such as to permit the instrument to be used in any position with respect to gravity. The possibility of breaking suspensions under severe shock as well as the side or vertical motion that might introduce rubbing friction had to be overcome.

The over-all size of the taut-band suspension assembly had to be proportioned in comparison to the pivot-jewel-spring assembly, so as to occupy approximately the same space in the instrument mechanism.

Element Sag Equivalent to Pivot-Jewel Side Play

One of the first important design problems was in respect to the amount of element sag. The amount of sag should not exceed the magnitude of side play in a pivot-jewel-type bearing instrument. The smaller this amount the less chance the moving element has of striking any stationary part of the instrument mechanism. Permissible sag had to be limited to allow the instrument to be used in any position with respect to gravity. The band suspensions are taut on either side of the element to the extent that they hold it on rotating center to within three-thousandths of an inch. Mathematically the limits of sag were calculated as follows (Fig. 2):

sine $A = S/L = P/2F$ or

$S = PL/2F$

where

S = the sag

P = element weight

L = length of the taut band

F = tautness or pull on the band

To obtain a small sag both the element weight P and the suspension band length L should be small and the tautness F large. The tautness force should be limited in practice to give a unit tensile stress in the band of approximately 50% of the yield of the band material. The unit tensile stress may vary in extremes from 40–60% of the yield point. This permits use of a given tension spring for slightly different band sizes to match various element torques.

In the new Westinghouse type *KX-241* 250-degree instrument (Fig. 3) the element weight is one gram, the band tautness is 100 grams, and the band length is 0.4 inch. Thus, the sag is $S = 0.4/200 = 0.002$ inch. The total side play is twice the sag or 0.004 inch, or the same order of magnitude as in pivot-jewel type-bearings.

Bumper Stops Solve Problems of Breakage Under Severe Shock and Side or Vertical Motion

The bumper and bumper stops shown in Fig. 1 overcome the possibility of breaking suspensions under severe shock, and relieve overstrains from side or vertical motions of the instrument without introducing any friction.

Zero Set

When the instrument pointer rotates up scale 250 degrees, the twisting end of

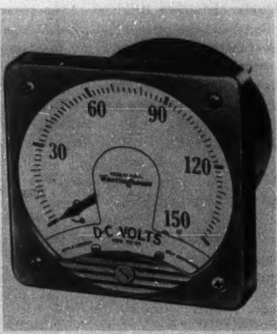


Fig. 3. Permanent magnet moving coil, 250-degree indicating instrument containing the new taut-band suspension element (Westinghouse type KX-241)

each suspension band also turns through 250 degrees with respect to its stationary end. The twisting of the suspension bands in this manner sets up a torsional stress in the band section. If the torsional stress were allowed to exceed the proportional limit of the suspension material, zero set would result. (Zero set is the deflection in degrees that an instrument pointer fails to return to zero when de-energized.) The maximum fiber stress in the cross section of the suspension band, in terms of maximum allowable twist, can be calculated by the following equation:[4]

$$f_s = Gt\theta/L$$

where

f_s = fiber stress, pounds per square inch
G = torsional modulus of the suspension band material, pounds per square inch
t = thickness of the band in inches
θ = twist angle (in the 250-degree instrument it is 4.36 radians)
L = length of the band (in the type *KX-241* element $L = 0.4$ of an inch)

With a given material and length, the maximum band thickness permissible is reached when the fiber stress f_s, in pounds and square inch, reaches the proportional limit of the band material. For no zero set it is then possible to determine a working thickness of a value less than that which will give the fiber stress its proportional limits. It will be noticed that the stress equation does not require consideration of the bandwidth. The bandwidth, therefore, can be selected to give the desired unit tensile strength, element sag, and torque. The fact that the bandwidth is independent of the torsional fiber stress is the reason for using ribbon stock instead of round wire. Not only is the bandwidth selected in terms of the desired tensile strength, sag, and torque, but it also is selected in terms of the desired current capacity and electrical resistance. The width also may be selected for other design reasons. For the *KX-241* instrument, band dimensions have been chosen to give negligible zero set. As tested in the laboratory under 20 to 1 magnification of the pointer and dial, a zero set of only 0.1 degree (0.04% of full scale) was found on the initial 16-hour full-scale deflection. A second full-scale deflection for 48 hours gave no additional zero set that could be detected with this magnification.

A more rigorous test was devised by putting 100 times full-scale current through the element holding the pointer against the full-scale stop. After 12 hours, a zero set of 0.3 degree was found; after 72 hours no further change in zero set was found just as with one milliampere through the element.

Torque Equation

The torque equation developed by Dr. S. Timoshenko[4] is applicable to an instrument element supported at either end by a taut-band suspension. The total torque equation is

$$Tq = 2[(bGWt^3\theta/L) + (TW^2\theta/12L) + (EtW^5\theta^3/360L)]$$

where

b = factor that varies directly with W/t
W = bandwidth, inches
T = band tension, grams, or the unit loading times cross-sectional area of the band
E = Young modulus, pounds per square inch

and other terms are as defined in the previous equations

With narrow bands, the major portion of the torque equation is in the twist component, or the first term of the equation. As the band gets wider the component part, the second term, of the equation becomes proportionally greater. The third term of the equation can be neglected for W/t ratios less than 20. This torque component is due to the elongation of the band edges and the shortening of the band center as the band twists. Where the ratio of bandwidth to thickness ratio exceeds 25, an increasingly nonlinear torque results, producing a nonlinear scale. A good design ratio for W/t is 10 to 1, or for convenience it may vary from 8 to 1 to 15 to 1 and linear torque is maintained.

Torque-to-Weight Ratio Is Not a Factor

The torque-to-weight ratio is frequently used to express the "factor of merit" in pivot-jewel instruments. This ratio indicates the relative ability of the instrument to overcome its bearing friction and give good response action. Of course, it is recognized that torque-to-weight ratio alone does not tell the whole story. The inevitable instrument friction also depends on the pivot and jewel material, dimensions, and workmanship, and on the kind of use the instrument gets. Since the suspension element has no friction surfaces, torque-to-weight ratio is not a factor in determining the friction merits of the instrument.

Torque-to-Inertia Is a Factor

Torque-to-inertia ratio is a more indicative factor of merit for the suspension-type element. That is, to obtain optimum response in an instrument mechanism the suspension element must have a relatively high ratio of torque-to-inertia. Its response time is a function of the damping and the free natural period of the element

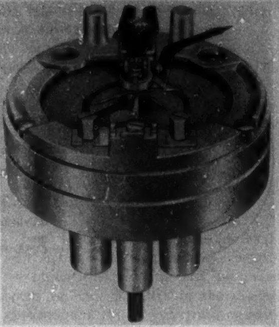


Fig. 4. Instrument mechanism removed from case showing the taut-band suspension element and magnet assembly

$$T_n = 2\pi\sqrt{I/U}$$

where I is the moment of inertia and U the torsional component. The equation shows that the response time varies directly as the square root of its moment of inertia I, and inversely as the square root of the torsional component U. By tests it was determined that good response time is 0.65 to 0.8 of T_n, for a damping factor of 100 to 400. This damping factor spread is for a pointer overshoot of 1 to 0.25% of the scale length respectively. Damping factors of less than 100 give a sudden increase in response time. On the other hand with damping factors greater than 400, the response time increases more slowly. For the optimum response, in accordance with the American Standards,[5] the instrument damping factor should be kept between 100 and 400.

Universal Element

With the type *KX-241* element a one-milliampere full-scale moving-element design is used universally for all ranges of ammeters and voltmeters. The weight of the moving element is one gram. The band suspension is sufficiently strong to permanently maintain a tension of 100 grams. The band is thin enough to twist through 250 degrees without zero set but thick enough to give a response time of $1\frac{1}{4}$ seconds. The resistance of the moving coil and suspension, with the necessary swamping resistance, is well below the limits required for application

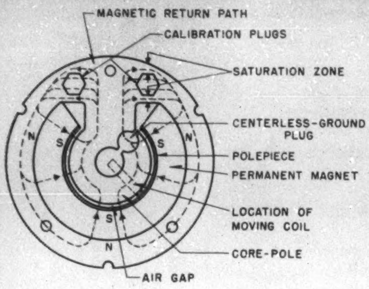


Fig. 5 (left). Magnetic circuit

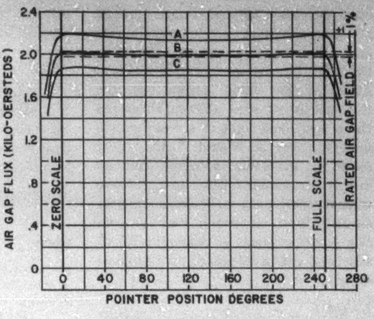


Fig. 6 (right). Curve showing adjustment range of the air-gap field

A—Maximum
B—Nominal
C—Minimum

with 50-millivolt shunts, and with shunt leads up to $1^1/_2$ ohms.

To obtain the optimum weight and inertia in the moving element, all parts are made of aluminum, including the moving coil winding. The moving element is used universally on all voltmeters with full-scale ranges from 50 millivolts to 800 volts, and on all ammeters with full-scale ranges from 1 milliampere to 50 amperes and higher. For lower ranges than these, where the current through the moving coil is the measured value, more turns of smaller diameter wire are used.

SELF-CENTERING TORSION HEAD ASSEMBLY

The complete moving element is self-centering, as shown in Figs. 1 and 4. The casting with the screw thread in the bridge is centered on the magnet circumference for air-gap alignment. The bumper stop screw which mounts in the bridge has a shoulder that locates the tension spring. The form heat-treated beryllium copper tension spring, when compressed to its specified position, centers the suspension band through the bumper stop screw. The other end of the suspension band is centrally attached to the band guide and anchor. The two element bumpers are cemented to the coil bracket with fixtures which align the bumpers with the axes of the suspension bands. The pointer and cross-arm assembly is staked to the top element bumpers. Conventional balance weights are used, and balancing is accomplished in the same manner as on a pivot-jewel-type instrument.

By compressing the tension springs to a specified point, the 0.4-inch suspension band lengths are maintained for accurate torque control. The tension strengths of the springs for each instrument are matched. The dimensions of the suspension band are held to a close tolerance. The band width is held to plus or minus one ten thousandth, and thickness to plus or minus five millionths of an inch. Instrument torque control on the production line therefore is held to extremely close limits.

PERMANENT MAGNET CIRCUIT

The permanent magnet circuit (Fig. 5) consists essentially of a strong Alnico magnet, a pole piece, and a circular lamination stamping of high-permeability high-saturation steel. The permanent magnet is a three-quarter circular ring of rectangular cross-section centerless ground for dimensional accuracy. The stamping is held to close dimensional tolerance and assembles to the magnet with a close fit. The round core section of the lamination extends into the magnet and forms a uniformly circular air gap. The magnetizing is radial which produces a field strength of 2,000 oersteds in the air gap.

One side of the element moving coil rotates in the air gap with ample clearance. The other side of the coil centers through the core opening along the rotating center of the suspension. The coil is assembled through an opening in the ring-shaped core. After the coil is assembled, the opening in the core ring, is closed with a centerless ground steel plug. With this plug in place, a balanced flux path is formed. This path carries the flux from the magnet across the air gap to the core, around to the outer section of the lamination and back to the magnet.

SERIES ADJUSTMENT

The strength of the air-gap field is controlled by a series adjuster in the following manner. In the outer section of the lamination two hexagon holes are placed in a narrow zone of the flux path. The material left in these restricted zones saturates unless similarly shaped plugs are inserted in the hexagon holes. By positioning these plugs the strength of the air-gap field is changed.

This series control is unique in that the air-gap field nearest the adjusted hexagon plug is affected the most. For instance, the scale ends of a zero-center instrument can be equalized and calibrated with these adjusters. On left-zero instruments, the full scale is calibrated by using the left plug and midscale by using the right plug.

To assure a uniform scale between the scale ends and center, a special shape pole piece is attached to the inner face of the magnet. The pole-width taper increases at a predetermined rate from the center to either end of the pole face in a manner to give a uniform scale or pointer deflection. An indication of this linearity is shown in Fig. 6.

Both of the foregoing methods of air-gap field control and adjustments are new features, used for the first time in this new instrument.

ASSEMBLY METHOD

As a subassembly of the instrument the pole piece is attached to the magnet, the magnet to the lamination stack, and all the laminations together by an epoxy resin binder. This means of assembly has been proved adequate for the high-shock requirements. As an example, the total shearing force required to push the magnet from the lamination assembly is 4 tons.

FRICTION

The advantages in eliminating pivot-jewel friction from indicating instruments cannot be overemphasized. The end results are greater sensitivity, increased damping, increased resistance to shock and vibration effect, and reduced maintenance.

Also, by eliminating friction the long-life accuracy rating of an instrument is improved by the amount of the potential friction error that might otherwise accumulate. Although the initial error due

Table I. Comparison of Performance of Taut-Band Instrument

Item	American Standards Requirements	Military Requirements	Type KX-*241*
Rated accuracy, % of full scale	1	1.5	1
Scale length, inches	6.8	6.5	7.1
Damping factor, ratio of full scale/overshoot	5 min	5 min	100–400
Response time, seconds	2.5 max		1.2–1.4
Sensitivity, full-scale milliamperes			
Ammeters	25 max	not listed	1
Voltmeters	11 max	not listed	1
External temperature influence, % change for 10 C.			
Ammeters	1 max	1 max	0.81
Voltmeters	1 max	1 max	inperceptible
External field influence, oersteds	5 for 3%	5 for 1%	225 for 1%
Maximum shunt leads, ohms			
Ammeters	not listed	not listed	1.5
Safe voltage-to-ground, peak value	1,100	not listed	1,385
Self-contained ratings			
Amperes	50	50	50
Volts	800	800	800
Ground tests, volts a-c effective	2,600	2,600	2,600
Shock tests, 1,500 times gravity	None	5% max	2–4%
Vibration			
% error	1	1	none
% error ¦friction	2	2	none

to friction in conventional instruments is small, it will increase with time. On suspension instruments the accuracy rating will remain stable throughout their prolonged service.

LONG LIFE AND LOW MAINTENANCE

With no pivots and bearings to contend with, the life of the instrument is extended beyond any now in use. Under accelerated life tests, the suspension instrument has been subjected to millions of operations without a perceptible change in calibration or change in zero setting. Although these particular test data have been obtained under laboratory conditions, it is sufficient to show that instrument repair and maintenance will be decreased considerably. There is no wear found in the taut-band suspension element even after 8 million zero-to-full-scale operations, and these tests are continuing with no indication that failure will occur. A conventional pivot-jewel-type instrument will require the replacement of the pivots before the first million of such operations.

SHOCK AND VIBRATION

With the conventional instrument the allowable amount of shock or vibration is determined by the durability of the pivot in conjunction with the jewel, and the amount of wear that will occur. The suspension type eliminates these problems. The ruggedness and extended life not only will provide for greater serviceability on present instrument applications, but also will open up new applications for electrical indicating instruments where severe shock and vibration have previously been limiting factors.

LOW ENERGY

Because of the low torque feasible with the suspension instrument, the energy required to operate the instrument is also very low. For example, a conventional 250-degree 1-milliampere instrument requires 0.18 milliwatt for full-scale indication. The suspension instrument, of the same rating, requires only 0.013 milliwatt. In terms of resistance, the conventional instrument of the 1 milliampere range has a resistance of 180 ohms, while the suspension instrument has only 13 ohms.

OVERLOAD

Even though the band is physically very small, and its resistivity comparatively high, it will take severe overloads; for example, at 150 times full-scale overload no damage is done to either the band or the moving coil. At this 15,000% overload the moving coil is required to dissipate less than 0.25 watt which is well within its continuous rating.

For a suddenly applied overload of several thousand per cent the instrument element rotates upscale at a rate directly proportional to the current applied. A current of 150 times full scale will produce a force of 150 times normal full-scale torque. When applying this high overload to a standard production sample of the type *X-241* instrument, allowing element to bump full-scale stops, no error or out of balance can be detected.

VIBRATION AND SHOCK

The Westinghouse type *KX-241* suspension instrument has been tested under the military vibration specification *MIL-Std-167*, simulated shipping tests to meet the requirements of the National Safety Transit Committee, and on various actual shipping conditions from the East Coast to the West Coast by truck, and in all cases the instruments maintained their initial accuracy of 1%. In addition they have been successfully tested under the Navy high-shock specification *MIL-M-16034.*[7]

Conclusions

This taut-band suspension development not only solves many major instrumentation problems of the past, but it also opens up new frontiers in the design, application, and use of electrical measuring instruments for the future.

Performance of the type *KX-241* compared with the requirements of the American Standard *C39.1*,[5] and military specifications[6,7] for 250-degree normal-scale 4$^1/_2$-inch rectangular switchboard d-c electric indicating instruments, see Table I.

References

1. ELECTRICAL MEASUREMENTS (book), **Frank A. Laws.** McGraw-Hill Book Company, Inc., New York, N. Y., 1938, p. 26.

2. ELECTRICAL MEASURING INSTRUMENTS—PART I, (book), **C. V. Drysdale, A. C. Jolley, G. F. Tagg.** John Wiley & Sons, Inc., New York, N. Y., 1952, pp. 3, 110–13.

3. TAUT-STRIP SUSPENSION IN INDUSTRIAL ELECTRICAL INSTRUMENTS, **Fritz Weingartner.** *Siemens Review*, Erlanger, Germany, Jan. 1956, pp. 14–20.

4. STRENGTH OF MATERIALS—Part II, **S. Timoshenko.** D. Van Nostrand Company, Inc., Princeton, N. J., third edition, pp. 288–91.

5. ELECTRIC INDICATING INSTRUMENTS (PANEL, SWITCHBOARD AND PORTABLE INSTRUMENTS). *ASA C39.1*, American Standards Association, New York, N. Y., Aug. 1955.

6. Type C, Mil S-901, *Mil. Std. 167*, Military Standard.

7. ELECTRICAL-INDICATING METERS (SWITCHBOARD AND PORTABLE TYPE). *Mil-M-16034A*, Military Specification.

◆

Discussion

H. P. Walker (Bureau of Ships, Washington, D. C.): The Bureau of Ships is extremely interested in the taut-band suspension for switchboard meters on the basis of improved operation under conditions of class high-shock and vibration. Consequently, Government laboratory tests were conducted on Westinghouse *KX241* 250-degree scale d-c switchboard milliammeters, to determine compliance with specification *MIL-M-16034A*.

Under exploratory vibration tests per *MIL-STD-167* using a rigid (1/2-inch steel plate) mounting, no zero shift or damage occurred to the meters tested. Total excursion was 2.6% of full scale value at 32 cps (cycles per second) and maximum pointer drift was 6.0% of full-scale value at 31 cps side-to-side motion.

When a 1/8-inch steel-plate mounting was used the two meters tested under side-to-side vibratory motion failed at 31 cps by breaking of the taut bands.

Two additional meters were subjected to class high shock. One meter failed at the 2,000-foot-pound back blow. Failure was caused by the cross-arm pointer assembly becoming disengaged from its supporting tower. Another meter also failed under humidity testing by disengagement of the pointer assembly.

It is urged that additional development work be applied to the taut-band suspension system to provide for a larger safety factor in the band design and a more positive method of securing the pointer.

V. S. Thomander and **R. C. MacIndoe:** The vibration tests performed on these instruments in accordance with *MIL-STD-167* resulted in no damage or change in zero. This agrees with the results reported in the second paragraph of Mr. Walker's discussion. In addition, we found the instruments to be within the limits given in *MIL-M-16034A* which includes changes in calibration due to vibration, and there was no friction effect. The instruments were attached to such rigid mounting brackets that the vibration experienced by the instrument was identical to that of the vibration machine and equal to the values specified in *MIL-STD-167*.

When the instruments are mounted on a less rigid mounting such as the $^1/_8$-inch steel plate referred to by Mr. Walker, the instrument experiences resonant amplification and excitation of other components of vibrations which are unknown but obviously are far in excess of the values in the specification. The specification contains no reference to the exact method of mounting or any mention of the use of a 1/8-inch steel plate. Because of this it is our interpretation of the specification that the values of frequency and amplitude given are the only ones that the instrument is to be subjected to. This was done successfully, as indicated by Mr. Walker's opening comments.

Since the tests referred to in Mr. Walker's discussion were performed, considerable additional improvements have been made in the strength of the joint that failed in both the shock and humidity tests. This was accomplished by changes in the formulation of the epoxy resin used and in the mechanical features of the parts in question. We have tested many instruments employing these improvements, and in all cases, they have successfully withstood both the humidity test and shock test specified in *MIL-M-16034A*.

These tests were performed at Westinghouse locations.

A Wide-Range Volt-Ampere Converter for Current and Voltage Measurements

F. L. HERMACH
MEMBER AIEE

E. S. WILLIAMS
NONMEMBER AIEE

TECHNOLOGICAL developments have brought increasing needs for convenient, portable instruments for measuring current and voltage with high accuracy over a wide range of frequencies. A "universal" volt-ampere (VA) converter has been developed for making such measurements to 0.05% with alternating current at frequencies from 5 cps (cycles per second) to 50 kc and with direct current. The new model is based on an earlier, much more limited prototype converter,[1] and, like it, is used with a null-type d-c potentiometer. With this single, self-contained, portable, 44-range instrument, the 1.5-volt range of a suitable potentiometer can be used to make a-c (rms) and d-c measurements from 7.5 milliamperes (ma) to 20 amperes and from 0.5 volt to 600 volts. With a more sensitive external galvanometer, the instrument can also be used as an a-c–d-c transfer standard for a-c–d-c difference measurements to 0.02% over the same frequency range.

Method

For a-c measurements, a single thermal converter (often called a thermoelement or thermocouple) serves as an a-c–d-c transfer standard. Its heater is connected in series with an a-c resistor for voltage measurements and in parallel with an a-c shunt for current measurements, as shown in Fig. 1. The electromotive force (emf) of the converter is manually balanced by the voltage from a built-in d-c source, with a galvanometer as the indicator (not shown in the figure). The heater of the converter is then switched to a second d-c source, which is adjusted to rebalance the galvanometer. The voltage across a portion of this circuit is then measured with the potentiometer and multiplied by a simple number to obtain the alternating voltage or current. For d-c measurements, the same multiplier is connected as a volt box and the same shunts are used, without the thermal converter.

From Fig. 1(A), with the thermal converter connected to R_m, $V_a=I_a(R_m+R_h)$. With the converter connected to the d-c circuit, the current through the heater is $I_h=I_dR_m'/(R_h+R_m')=V_pR_m'/R_p(R_h+R_m')$. If the a-c–d-c difference of the converter is negligible, $I_a=I_h$, and if $R_m'=R_m$, $V_a=V_pR_m/R_p$. The a-c voltage to be measured is thus the product of the measured potentiometer voltage and a ratio of resistances, and is independent of both the emf-versus-current characteristic and the heater resistance of the converter.

From Fig. 1(B), with the converter connected to R_s, the current through the heater is $I_h=I_aR_s/(R_h+R_s)$. With the converter in the d-c circuit, $V_p=I_d(R_h+R_s')$. If $I_h=I_d$ and $R_s'=R_s$, and if the a-c impedance of the heater is equal to its d-c resistance, $I_a=V_p/R_s$, again desirably independent of the emf-versus-current characteristic and the resistance of the converter.

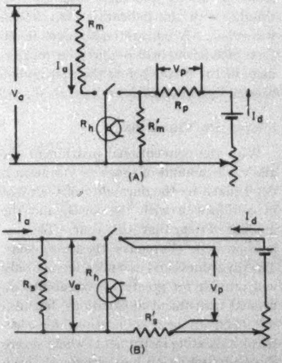


Fig. 1. Schematic diagram of VA converter

Paper 59-161, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 1, 1957; made available for printing December 1, 1958.

F. L. HERMACH and E. S. WILLIAMS are with the National Bureau of Standards, U. S. Department of Commerce, Washington, D. C.

Vol.
Pt. 1
1959

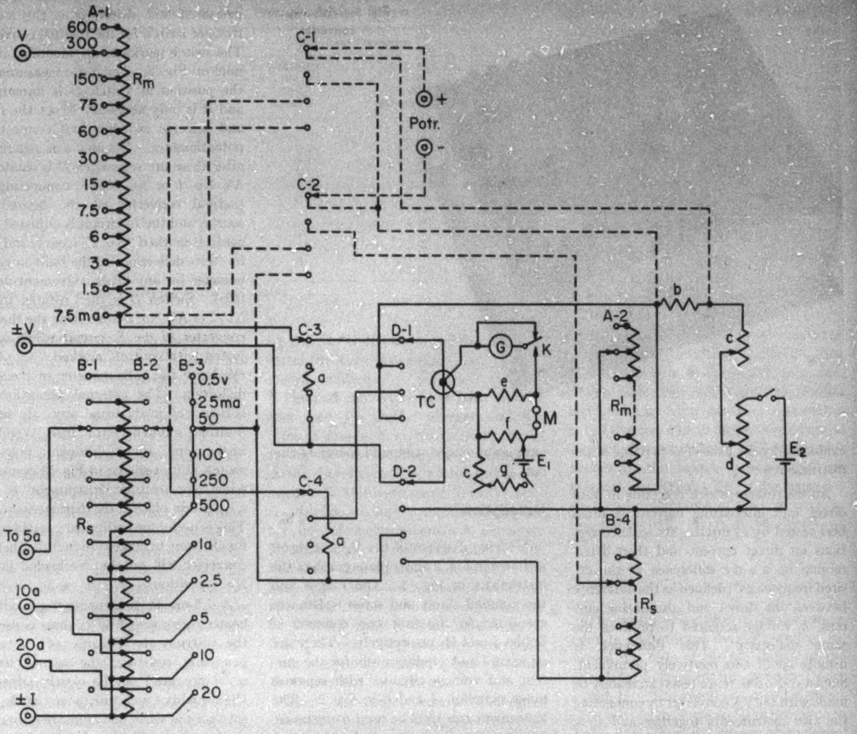


Fig. 2. Wiring diagram of VA converter. Switch-position markings from top to bottom on diagram: switch C, AV, AC, DV, DC; switch D, AV No. 1, AV No. 2, AC No. 1, AC No. 2. Dotted lines signify potential leads except at A-2 and B-3 where they indicate intermediate switch positions are not shown

A-1, etc. = pole 1 of switch A
A, B = voltage and current range switches
C, D = selector and transfer switches
TC = thermal converter, 7.5 ma
G = galvanometer, 0.3 microampere per centimeter
K = galvanometer key
a = 66-²/₃ ohms, manganin
b = (Rp) 200 ohms, manganin
c = 10 ohms, 3-turn voltage divider
d = 1,000 ohms, 3-turn voltage divider
e = 10 ohms manganin
f = 100 ohms, 3-turn voltage divider
g = 700 ohms, manganin
E_1 = 1.5 volts, no. 6 dry cell
E_2 = 3 volts, no. 6 dry cells
M = binding posts for external milliammeter

For direct voltage measurements, the heater of the thermal converter is disconnected, the multiplier, R_m, is connected directly to the voltage to be measured, and a 1.5-volt section of it is connected to the potentiometer. If this section is designated as R_a, $V_{dc}=V_pR_m/R_a$, and if $R_a=R_p$ the nominal multiplying factor, $F_e=R_m/R_a$, for the volt box thus formed is the same as that for the corresponding a-c voltage range. For d-c measurements the potential terminals of the shunt R_s are connected directly to the potentiometer, so that $I_{dc}=V_p/R_s$, and again the nominal multiplying factor, $F_i=1/R_s$, is the same as that for the corresponding a-c range.

If the resistors R_p, R_s, R_m, R_m', and R_s' have small proportional corrections of a_p, a_s, a_m, a_m', and a_s' respectively, more detailed analysis shows that, to a sufficient approximation,

$$V_a=\frac{V_pR_m}{R_p}\left[1-a_p+\frac{a_m+a_m'R_h/R_m}{1+R_h/R_m}\right]=V_pF_e(1+v) \quad (1)$$

and

$$I_a=\frac{V_p}{R_s}\left[1-\frac{a_s+a_s'R_s/R_h}{1+R_s/R_h}\right]=V_pF_i\,(1+w) \quad (2)$$

where the capital letters now signify the nominal values. For the useful ranges, $R_m>R_h>R_s$, so that the accuracy required of the compensating resistors R_m' and R_s' is roughly inversely proportional to the current or voltage range, while essentially full accuracy is required of R_m, R_s, and R_p. In practice the over-all correction factors, v and w, are directly

Fig. 3. Volt-ampere converter

evaluated by d-c tests of each range of the instruments.

An instrument which responds to both direct and alternating current is often best tested by evaluating its scale corrections on direct current and then determining its a-c–d-c difference at the desired frequencies[2] (defined as the difference between the direct and alternating current or voltage required to produce the same deflection). This difference is usually small and relatively permanent. Such a-c–d-c difference tests can readily be made with the VA converter by connecting the two instruments together and to a source of the desired voltage or current, making a preliminary setting of the test instrument, and then balancing the emf of the thermal converter, *E*, to give zero deflection of the galvanometer. The instruments are then successively switched to a-c, d-c, reversed d-c, and a-c sources, which are adjusted to give the same deflection of the test instrument, and the resulting galvanometer deflections are observed, without rebalancing. The percentage a-c–d-c difference of the test instrument, *T*, is computed as $T = \Delta D/KE$ where ΔD is the difference between the mean of the two a-c and the two d-c galvanometer deflections, and *K* is a sensitivity factor. This factor is readily determined by observing the changes in galvanometer deflection resulting from small measured differences in applied direct current or voltage at several different levels of heater current, and plotting *K* versus *E*. Tests show that only two such curves are necessary, one of which applies to all current ranges and the other to all voltage ranges. (They differ because of the dependence of heater resistance upon temperature and therefore current.)

Description

A wiring diagram of the VA converter is shown in Fig. 2 and a photograph of the instrument in Fig. 3. The ranges and the nominal shunt and series resistances are listed in the first two columns of Tables I and II, respectively. There are separate input binding posts for the current and voltage circuits, with separate range switches, *A* and *B* in Fig. 2. The instrument can thus be used to measure, in succession, the voltage and current in a d-c or a-c circuit without reconnections. It may also be useful for wattmeter testing on direct current or at unity power factor on alternating current (more strictly at the power factor of maximum deflection of a wattmeter) by the familiar "phantom loading method."[3]

Switch *C* selects the quantity to be measured and switch *D* is the a-c–d-c transfer switch for the thermal converter. The switch markings are identified in the note on Fig. 2. For a d-c measurement, the position of switch *D* is immaterial, and it is only necessary to set the range and selector switches and connect the potentiometer. For an a-c measurement, after these are set switch *D* is rotated to AV No. *1* or AC No. *1*, connecting the thermal converter to the desired a-c source, and the E_1 circuit is adjusted with controls marked "No. *1*" (coarse and fine) for zero deflection of the built-in galvanometer (or any other convenient deflection). Switch *D* is then rotated to AV No. *2* or AC No. *2* connecting the thermal converter to the E_2 circuit which is adjusted with controls marked "No. *2*" to rebalance the galvanometer to the same deflection. The external potentiometer is then read in the usual way. In switch positions marked "Rect. Inst.," (rectifier instrument) two additional poles of switch *D* (not shown in Fig. 2) connect a panel-type rectifier instrument to the a-c input in place of the thermal converter. This is useful for preliminary settings and for checking to make sure that the thermal converter will not be overloaded in the No. *1* position.

A 7.5-ma thermal converter having a heater resistance of $66^2/_3$ ohms is used in the instrument. Because of the compensating resistors, the actual heater resistance need not be closely adjusted. High-quality a-c resistance cards are used for the series and shunt resistors of 1 ohm or more. Shunts of lower resistance, for the higher current ranges, were specially made in our laboratories of bifilar strip, ranging from 2 to 20 mils in thickness, of nickel-chromium-aluminum-iron alloy, with insulation of 1-mil polyester film. The dimensions were chosen as a compromise between the requirements of large area for cooling and of thin strip for

Table I. Results of D-C Tests of Current Ranges

Current Range, Amperes	R_s, Ohms	Per-Cent Correction: D-C Ranges, at 9 Months*, Rated I	Per-Cent Correction: D-C Ranges, at 9 Months*, 20% Rated I	Per-Cent Correction: A-C Ranges, at Rated I, At 6 Months*	Per-Cent Correction: A-C Ranges, at Rated I, At 18 Months*
20	0.025	+0.05	−0.07	+0.03	+0.05
10	0.05	−0.01	−0.09	+0.01	+0.02
5	0.1	−0.02	−0.01	+0.04	+0.03
2.5	0.2	+0.02	+0.03	+0.03	+0.05
1	0.5	0.00	+0.03	+0.03	+0.05
500 ma	1	−0.09	−0.08	−0.07	−0.02†
250 ma	2	−0.04	−0.03	−0.02	0.00
100 ma	5	−0.01	−0.01	−0.01	0.00
50 ma	10	−0.04	−0.04	−0.04	−0.03
25 ma	20	−0.03	−0.03	−0.02	−0.02
7.5 ma	None			+0.01	

* Number of months after completion of the instrument.

† Readjusted before this test; adjustment affects lower-current ranges.

Table II. Results of D-C Tests of Voltage Ranges

Voltage Range, Volts	R_m, Kilohms	Per-Cent Correction			
		D-C Ranges, at 9 Months*		A-C Ranges, at Rated V	
		Rated V	20% Rated V	At 6 Months*	At 18 Months*
600	80	+0.01	+0.01		+0.02
300	40	+0.00	0.00	+0.02	+0.01
150	20	+0.00	0.00	+0.02	+0.01
75	10	−0.01	−0.01	+0.01	0.00
60	8	−0.01	−0.01	0.00	0.00
30	4	0.00	0.00	+0.01	+0.02
15	2	0.00		+0.02	+0.02
7.5	1	−0.02		−0.02	−0.02
6	0.8	−0.01		−0.02	0.00
3	0.4	0.00		−0.03	−0.03
1.5	0.2	0.00		−0.01	−0.03
0.5†	None			+0.02	

* Number of months after completion of the instruments.
† Alternating-current range only.

minimum reactance, using inductance and skin-effect formulas developed by Silsbee.[4] The compensating resistors, which carry only small direct current, were chosen to meet the accuracy requirements dictated by equations 1 and 2.

A built-in taut-suspension light-beam galvanometer with 50-millimeter (mm) scale provides a deflection of 2 mm for 0.1% change in input current or voltage at rated heater current. This is adequate for the zero settings required in a-c tests, but for a-c–d-c transfer tests to 0.02%, in which differences in deflection must be read, a more sensitive external box-type galvanometer with a longer scale is connected in place of the internal galvanometer. For these tests an external 1.5-ma d-c milliammeter is also connected at a pair of binding posts marked *M* in Fig. 2, to measure the battery current through the 10-ohm resistor, *e*, and therefore, at balance, the emf, *E*, of the thermocouple.

The VA converter has over-all dimensions of 17x14x9 inches and weighs 30 pounds.

Precautions

A number of factors must be evaluated in the design and construction of an instrument such as this. Some are commonly encountered in the design of any equipment of high accuracy such as a d-c potentiometer, but are accentuated by the high voltage used in some of the circuits. These are: 1. insulation, contact, and lead resistance of the wiring, switches, and other components; 2. self-heating and ambient temperature effects in the shunt and series resistors; and 3. the stability of these resistors. In addition, because of the unusual frequency range in the instrument, the effects of residual reactances, of the a-c resistors, associated circuits, and of the frequency influence of the thermal converter must be carefully considered. Still other factors are peculiar to the use of the thermal converter in this instrument and deserve special mention.

Thermal converters have been studied and used as highly accurate a-c–d-c transfer standards.[5] Normally the average for the two directions of current serves as the d-c reference in transfer measurements. In this instrument, however, only one direction is used for convenience and rapidity of measurement. A converter of small d-c reversal difference is thus desirable. Fortunately, converters of the bead type, with an insulating bead between the heater and thermocouple, can be selected with reversal differences of 0.02% or less. They are then completely interchangeable without requiring recalibration of the instrument.

Most low-range thermal converters have ambient temperature coefficients of 0.1 to 0.2% per degree centigrade (C), and should be thermally protected for precise transfer measurements. A cotton-filled Bakelite cylinder around the converter and the wooden case of the instrument provide a thermal time constant of 40 minutes, which is ample for normal laboratory use of this instrument. A wooden divider separates the shunts and multiplier from the converter and the d-c circuits to minimize self-heating changes.

After a sudden change in heater current, many bead-type thermal converters exhibit an exponential "warm-up" drift having a time constant of 1 to 2 minutes. This drift, an increase of current with time for the same emf, can amount to almost 0.1% for some converters following a change from zero to rated current. The percentage error caused by it is

$$d=\int_{t_1}^{t_2} S\,dt$$

where S is the percentage rate of change of heater current at constant emf and t_2 and t_1 are the elapsed times at which the No. *2* controls and the potentiometer are balanced, respectively. The drift rate did not exceed 0.04% per minute for any 7.5-ma converter tested to date, so that, since 1/2 minute is ordinarily sufficient time for these balances, the warm-up drift error should not exceed 0.02% even for the largest change in heater current. However, for best results, a warm-up period of at least 2 minutes should be allowed. After the initial change, the steady-state drift of the galvanometer in either the No. *1* or No. *2* positions of switch *D* is less than 0.003% per minute, with No. *6* dry cells as the internal d-c sources.

The a-c resistors permit some response of the thermal converter at frequencies up to 100 megacycles (mc). In areas of high television signal strengths, errors can result, since at these frequencies laboratory leads of ordinary lengths can approximate quarter-wave antennas. These difficulties were avoided in this instrument by a low-pass L-C filter with a cutoff frequency of 10 mc per second and a characteristic impedance of 67 ohms, connected to the heater terminals of the thermal converter. The filter changed the a-c–d-c difference of the converter by less than 0.04% [0.003 db (decibels)] at 100 kc.

Tests

The results of d-c tests of the d-c ranges (shunts and voltage divider), and of reversed d-c tests of the a-c ranges of the completed instrument are given in Tables I and II. The results of a-c–d-c difference tests from 20 cps to 50 kc are given in Table III. The d-c shunts were tested with an accurate Kelvin double bridge, and the volt box tested in comparison with the National Bureau of Standards

Table III. Results of A-C–D-C Difference Tests at Rated Current and Voltage

Range	Per Cent A-C–D-C Difference			
	20 Cps	20 Kc	30 Kc	50 Kc
600 volts	0.00	0.00	0.00	+0.03
300 volts		−0.01	−0.01	−0.02
150 volts	0.00		−0.01	−0.02
75 volts			0.00	−0.02
60 volts			0.00	−0.01
30 volts			0.00	0.00
15 volts			+0.01	0.00
7.5 volts			+0.01	0.00
6 volts			+0.01	+0.01
3 volts			+0.01	+0.01
1.5 volts	0.00		0.00	+0.01
0.5 volt	0.00		+0.02	+0.03
20 amperes		+0.01	−0.01	−0.02
10 amperes		−0.01	−0.01	−0.04
5 amperes	0.00	−0.02	−0.03	−0.04
2.5 amperes			0.00	+0.01
1 ampere			0.00	+0.04
500 ma	0.00		−0.01	−0.03
250 ma			−0.02	−0.01
100 ma			0.00	−0.01
50 ma			0.00	+0.01
25 ma	0.00		+0.01	+0.01
7.5 ma			−0.01	+0.01

(NBS) standard volt box. The reversed d-c tests were made with a potentiometer and auxiliary shunts and volt boxes. The a-c–d-c difference tests were made with the NBS standard thermal converters,[5] which have recently been evaluated down to 2 cps, and with a newly constructed transfer voltmeter element, which has been evaluated to 100 kc.

The results of the d-c tests show generally satisfactory performance, adjustment, and stability of the instrument, except for the 10- and 20-ampere ranges in which much larger shunts should have been used. Separate correction curves have been prepared for these ranges to show the change in resistance caused by the self-heating of the measured current.

The a-c–d-c differences, defined as $(I_{ac}-I_{dc})/I_{dc}$ or $(V_{ac}-V_{dc})/V_{dc}$ for the same thermocouple emf with alternating and direct current, are very small. They depend upon the ratios of small reactances to resistances in the instrument and may be expected to remain constant to the accuracy desired. Thus, unless the components are changed, future periodic tests of the instrument need normally be made with direct current only, with a-c–d-c checks made only occasionally. This greatly simplifies the standardization of the instrument.

Additional low-frequency tests of the 0.5-volt and 7.5-ma ranges showed that the a-c–d-c difference of the converter itself is not more than 0.01% at frequencies down to 5 cps. Both theory and experiment indicate that the difference should also be negligible on all other ranges of the instrument.

The small discrepancies in Tables I and II between the corrections for the corresponding d-c and a-c ranges are caused by the d-c reversal difference of the thermal converter, the difference between R_a and R_p, and the influence of the compensating resistors R_m' and R_s'. While these are small (the cause of the relatively large discrepancy [0.06%] between the corrections for the d-c and the a-c 5-ampere ranges is unknown), they are difficult to evaluate separately, so that future d-c tests must probably be made on both a-c and d-c ranges. In most laboratories a volt box and shunt box could be reserved as standards for these and other tests and could be sent to a suitably equipped standardizing laboratory periodically, or the instrument itself could be sent to such a laboratory.

Conclusions

This portable, VA converter can serve as an accessory to a null-type d-c potentiometer to make possible d-c and a-c measurements of very high accuracy over truly wide ranges of voltage, current, and frequency. Its accuracy depends primarily upon wire-wound resistors and strip shunts of excellent stability and low-temperature coefficient. After initial a-c–d-c difference tests of the instrument, routine periodic tests need ordinarily be made only with direct current, thus simplifying the standardization of the instrument.

The VA converter should be particularly suitable for testing other instruments, since all three types of tests, (d-c, a-c, and a-c–d-c transfer tests) may be made with it.

References

1. MULTIRANGE, AUDIOFREQUENCY THERMOCOUPLE INSTRUMENTS OF HIGH ACCURACY, F. L. Hermach, E. S. Williams. *Journal of Research*, National Bureau of Standards, Washington, D. C., vol. 52, 1954, p. 227.

2. THE TESTING OF ELECTRICAL INSTRUMENTS, F. L. Hermach. *Proceedings*, Instrument Society of America, Pittsburgh, Pa., vol. 8, 1953, Paper 53-2-3.

3. ELECTRICAL MEASUREMENTS (book), F. K. Harris. John Wiley & Sons, Inc., New York, N. Y. .1952, chap. 11.

4. NOTES ON THE DESIGN OF FOUR TERMINAL RESISTANCE STANDARDS FOR ALTERNATING CURRENT, F. B. Silsbee. *Journal of Research*, National Bureau of Standards, vol. 4, 1930, p. 73.

5. THERMAL CONVERTERS AS AC-DC TRANSFER STANDARDS FOR CURRENT AND VOLTAGE MEASUREMENTS AT AUDIO FREQUENCIES, F. L. Hermach. *Ibid.*, vol. 48 ,1952, p. 121.

Vol.
Pt. 1
1959

An Impedance Bridge for Surface Temperature Measurement

R. J. MOULY
ASSOCIATE MEMBER AIEE

TEMPERATURE MEASUREMENT and control is a basic requirement in many industrial processes. A common problem is the measurement of surface temperatures without contacting the object because it is inaccessible, moving, or because contacting might damage the surface finish. For example, this is the type of problem encountered when it is desired to control the surface temperature of metal molds or rolls in order to reproduce or maintain temperature conditions which are found to give a satisfactory product. The trend toward higher quality and uniformity of processed products results in increasingly closer temperature limits and correspondingly exacting requirements for the temperature-measuring instrument.

For these applications, the reliability of currently available instruments is often questionable. Some of the limitations of these instruments are listed.

1. Radiation pyrometers are subject to two important sources of error:
   a. The emissivity of highly polished surfaces is often uncertain and varying.
   b. An error results from the absorption of radiation by gases or liquids interposed between the measuring element and the surface whose temperature is to be measured.
2. Aspirating pyrometers are sometimes used, but their application is somewhat limited for moving surfaces such as rolls. They require direct access to the object whose temperature is to be measured and cannot be used whenever adverse ambient conditions exist.

An inductive method of temperature measurement that is free of the foregoing limitations was proposed in the early thirties by G. Keinath.[1] This technique pertains to surface temperature measurements of electrically conductive bodies by means of a special impedance bridge. When an inductive transducer is placed in the vicinity of the surface whose temperature will affect the eddy currents induced in the metal resulting in a change in the impedance of the transducer; this variation unbalances a bridge and is detected by a phase-sensing detector that eliminates the effect of minute distance variations between the transducer and the surface. This measurement method, however, is applicable only to the detection of small temperature deviations as the insensitivity of the output to distance variations is obtained

---

Paper 59-158, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 7, 1958; made available for printing December 5, 1958.

R. J. MOULY is with the Corning Glass Works, Corning, N. Y.

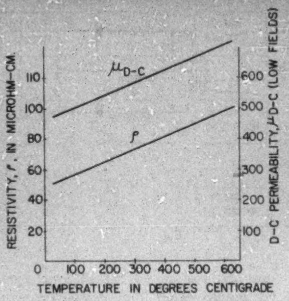


Fig. 1. Resistivity and d-c permeability versus temperature for magnetic stainless-steel type 420

within a relatively narrow temperature range around a preset value.

This paper presents an instrument that permits accurate emissivity-free temperature measurements over a wide range. The method is noncontacting, not even direct (optical) view is required, the speed of response is inherently high, and adverse ambient conditions do not affect the instrument.

TEMPERATURE MEASUREMENT PROBLEM

This instrument has been developed for a specific application: the measurement of the temperature of the glass-contacting surface of molds used in the pressing of glass objects. In some instances, it is desired to control the temperature of the glass-contacting surface very closely, within a few degrees centigrade (C). The measurement problem consists of detecting deviations of the surface temperature of the mold from a reference value.

The emissivity of the mold surface is known to vary with time, therefore, normal temperature measuring techniques may be inaccurate.

Preliminary Considerations

EFFECT OF TEMPERATURE ON THE ELECTRICAL AND MAGNETIC PROPERTIES OF THE METAL

The molds are made of magnetic stainless steel. The surface temperature, when measured, is of the order of 450 to 500 C. At these temperatures, well under the Curie point, the metal still retains its magnetic characteristics and both the permeability and resistivity steadily increase with temperature as shown in Fig. 1. The data on magnetic permeability are only indicative, as the permeability depends upon the flux density.

The metal is in a fully annealed condition and, as its spheroidized microstructure is extremely stable, its electrical properties do not change with time. These properties are therefore functions of temperature only.

EDDY CURRENTS IN MASSIVE CONDUCTIVE BODIES

The temperature measurement technique is based on the generation of eddy currents in the body whose temperature is to be measured. The material is subjected to a sinusoidally varying field and eddy currents are generated that depend on the electrical properties of the metal, resistivity and permeability, and the frequency of alternation of the flux.

If the field is applied parallel to the surface, its amplitude decreases exponentially from the surface to the interior as shown in Fig. 2 for different frequencies.[2] This is true if the thickness of the body is large as compared to the skin thickness which characterizes the penetration of the surface field into the interior and is defined as the distance s from the surface at which the field is $1/\epsilon$ times or 36.8% its surface value.

$$s=5{,}030/\sqrt{\mu f/p}, \text{ centimeters (cm)}$$

ρ = resistivity of material, ohm-cm
μ = permeability of material
f = frequency, cycles per second (cps)

In the particular case under consideration, $\rho=90\ 10^{-1}$ ohm-cm; $\mu=600$ (for low fields), f=200 cps; then $s=0.14$ cm (approximately 1/8 inch). At this frequency, the skin thickness is small compared to the thickness of the mold which is about two inches. Consequently, the eddy currents will be mainly dependent on the electrical and magnetic properties of the metal in the zone immediately beneath the surface of the mold and will be affected by the variations of these properties with temperature.

CHANGE OF INDUCTIVE AND RESISTIVE IMPEDANCES OF AN INDUCTOR PLACED AT A VARYING DISTANCE FROM A SHEET OF MAGNETIC STAINLESS STEEL

An inductor is characterized electrically by the two components of its impedance; its resistance R and its inductive reactance $2\pi fL=\omega L$. If an iron-cored coil, such as represented in Fig. 3, is placed with its axis normal to a thick plate of magnetic stainless steel at a distance d from the surface, its resistance and its inductive reactance will increase. Eddy currents and hyteresis losses will increase

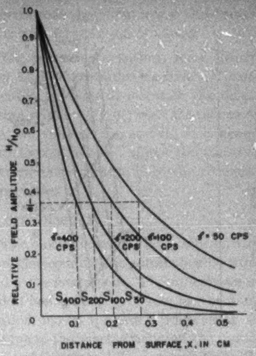


Fig. 2. Magnetic field strength distribution in a thick plate of uniform permeability and resistivity

$\mu=600$
$\rho=90\ 10^{-6}$ ohm-cm
f=50, 100, 200, 400 cps
$H/H_0=\epsilon^{-2\pi\sqrt{\mu f/\rho}(x)}$

$$s_f=\frac{1}{2\pi\sqrt{\mu f/\rho}}$$

= skin thickness at frequency f

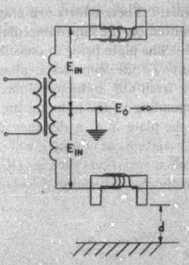


Fig. 3. Basic bridge circuit

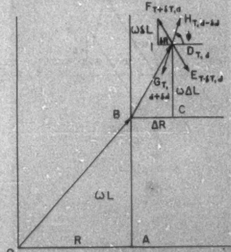


Fig. 4. Effect of temperature distance variations on probe impedance

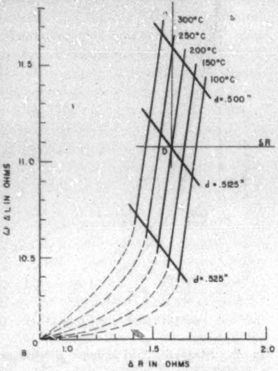


Fig. 5. Resistance reactance plot. Effect of distance and temperature variations

f=200 cps
$R_{initial}$=700 ohms
$\omega L_{initial}$=300 ohms

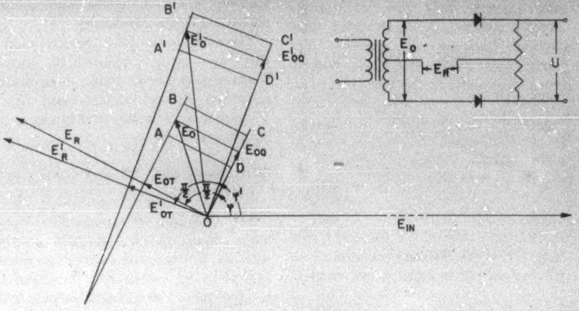


Fig. 7. Phasor diagram for distance modulated probe

the resistance by ΔR and the decrease of the reluctance of the magnetic circuit will result in an increase $\omega \Delta L$ of the inductive reactance. These effects are graphically represented on the impedance diagram in Fig. 4. The plate being at constant temperature T, the impedance phasor will change from OB, inductor alone, to OD corresponding to the inductor impedance when the plate is at a distance d. Distance variations δd around d will result in impedance variations such as DG or DH corresponding to distances $d+\delta d$ or $d-\delta d$

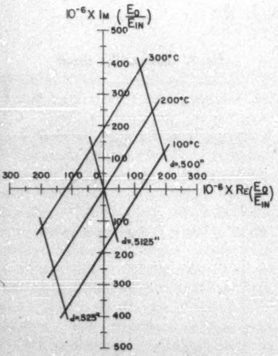


Fig. 6. Bridge unbalance voltage versus impedance variations

See Fig. 5 for f, $R_{initial}$, and $\omega L_{initial}$

EFFECT OF METAL TEMPERATURE CHANGES

If the distance is maintained constant and the temperature of the plate is raised from T to $T+\delta T$, its permeability and resistivity will increase resulting in an inductive reactance increase ($\omega\delta L$), IF in Fig. 4, and a decrease in the iron losses. The latter change appears as a resistance decrease δR, ID in Fig. 4. The impedance of the system will now be OF. Similarly, a plate temperature decrease will result in an impedance change DE. Hence, the direction of an impedance change on the impedance plane is characteristic of the physical origin of the change: distance or temperature.

RELATIVE SENSITIVITIES TO DISTANCE AND TEMPERATURE VARIATIONS—EXPERIMENTAL DATA

The preceding considerations were checked experimentally in order to evaluate quantitatively the order of magnitude of the changes.

Fig. 5 summarizes the results of measurements made at 200 cps on a system consisting of an inductor and a plate of stainless steel in the 100–300 C temperature range and at distances varying from 0.5 to 0.525 inch.

Within that range, the variations of the impedance with distance and temperature around its initial value, corresponding to $d=0.5125$ inch and $T=200$ C are well represented by the linear empirical equations:

$$\delta R = a_{11}\delta d + a_{12}\delta T$$

$$\omega\delta L = a_{21}\delta d + a_{22}\delta T$$

$a_{11} = -5.6$ ohm/inch; $a_{12} = -1.3 \ 10^{-6}$ ohm/C

$a_{21} = -37.2$ ohm/inch; $a_{22} = 1.75 \ 10^{-6}$ ohm/C

The sensitivity to distance variations is such that a distance variation of the order of 10^{-5} inch affects the impedance components as much as a plate temperature change of 1 C.

In an industrial process, minute distance changes of that order are normally unavoidable; it is not even generally possible to define the location of the surface with such accuracies. In order to attain precision in temperature measurement, the output of the instrument will have to be insensitive to minute distance variations between the sensing inductor and the surface whose temperature is being measured.

Instrument Design

MEASUREMENT FREQUENCY

It would seem that the optimum frequency for surface temperature measurement should be as high as possible, as high measurement frequencies are associated with high sensitivities and thin skin thicknesses; however, increasing the measurement frequency results in increased relative sensitivity to surface finish conditions. In the present case, the surface of the stainless-steel molds is plated with a thin layer of chrome whose thickness (maximum 0.0004 inch) and characteristics change with time. It is essential that the temperature indication of the instrument should not be affected by these changing conditions of the surface whose temperature is to be measured. This insensitivity is obtained by lowering the measurement frequency to a value at which the temperature error does not exceed 1 C on the instrument indications.

DESIGN BASIS FOR THE TEMPERATURE MEASURING BRIDGE

The instrumentation problem consists of designing circuitry that will discriminate between the changes of the inductor

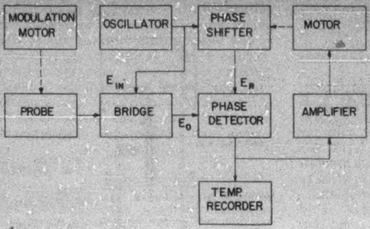


Fig. 8. Circuit for measuring metal plate surface temperature

impedance caused by surface temperature variations and those caused by fluctuations of the probe to surface spacing, and provide an output that is a function of the plate temperature only.

The adopted solution is basically an impedance comparison bridge followed by a phase detector. In an impedance bridge, such as represented in Fig. 3, if the voltages E_{in} across the windings are equal, the real and imaginary parts of the ratio of the unbalance voltage E_o to the input voltage E_{in} are

$$Re\left(\frac{E_o}{E_{in}}\right) \simeq \frac{1}{2}\,\frac{R\delta R+\omega^2 L\delta L}{R^2+L^2\omega^2}$$

$$Im\left(\frac{E_o}{E_{in}}\right) \simeq \frac{1}{2}\,\frac{R\omega\delta L-\omega L\delta R}{R^2+L^2\omega^2}$$

R and ωL being the resistance and the reactance of the probe in the presence of the plate at the initial distance and temperature for which the bridge is balanced. The graph, Fig. 6, shows the variations of the output voltage of the bridge for changes of the plate temperature T and the probe spacing d around their initial values $d=0.5125$ inch and $T=200$ C.

The unbalance voltage components expressed in terms of temperature and distance variations are

$$Re\left(\frac{E_o}{E_{in}}\right) \simeq b_{11}\delta d+b_{12}\delta T$$

$$Im\left(\frac{E_o}{E_{in}}\right) \simeq b_{21}\delta d+b_{22}\delta T$$

with

$$b_{11}=\frac{1}{2}\,\frac{Ra_{11}+L\omega a_{21}}{R^2+L^2\omega^2}$$

$$b_{21}=\frac{1}{2}\,\frac{Ra_{21}-L\omega a_{11}}{R^2+L^2\omega^2}$$

$$b_{12}=\frac{1}{2}\,\frac{Ra_{12}+L\omega a_{22}}{R^2+L^2\omega^2}$$

$$b_{22}=\frac{1}{2}\,\frac{Ra_{22}-L\omega a_{12}}{R^2+L^2\omega^2}$$

The surface temperature and spacing variations result in unbalance voltages of distinct phase.

Any unbalance voltage E_o resulting from simultaneous variations of the temperature and the distance can be resolved into two components E_{oT} and E_{oQ} (Fig. 7). E_{oT} is the component independent of distance and E_{oQ} is the component at right angle to it

$$\frac{E_{OT}}{E_{in}}=\frac{-b_{12}b_{21}+b_{11}b_{22}}{\sqrt{b_{11}^2+b_{21}^2}}\,\delta T$$

$$\frac{E_{OQ}}{E_{in}}=\sqrt{b_{11}^2+b_{21}^2}\,\delta d+\frac{b_{11}b_{12}+b_{21}b_{22}}{\sqrt{b_{11}^2+b_{21}^2}}\,\delta T$$

E_{OT} is temperature dependent only and is the voltage to be measured.

A phase detector, Fig. 7, will detect that component of the error voltage that is in phase with the reference voltage E_R. If the reference is in quadrature with E_{oQ} the d-c output U of the phase detector will be proportional to E_{QT} and depends upon δT only.

Continuous Adjustment of the Phase of the Reference Voltage

The preceding equations show clearly that E_{oQ} is temperature sensitive. Consequently, in order to maintain the quadrature relationship between the phases of the two voltages E_{oQ} and E_R, it is necessary to continuously adjust the phase of the reference voltage when the temperature varies.

Block Diagram

The block diagram, Fig. 8, schematizes how this condition is automatically satisfied in the instrument. The probe is vibrated at low frequency, 20 cps, by the modulation motor, in order sinusoidally to modulate the spacing between the inductor and the plate. The error voltage E_o, Fig. 7, oscillates between OA and OB and is amplitude and phase modulated. The phase of the reference voltage E_R is adjustable by the phase shifter whose position is controlled by a servo system.

As long as the reference voltage is not in quadrature with E_{oQ}, a 20-cps signal is present in the output of the phase detector. This a-c component is the error signal input of the servo system that adjusts the resolver position until the 20-cps component disappears from the output of the phase detector, at which time the reference voltage is in quadrature with E_{oQ} and the output of the phase detector is a function of δT only.

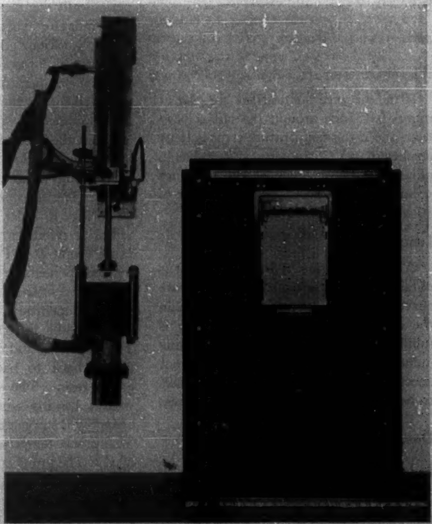


Fig. 9. Surface temperature measuring system showing inductive probe, left, power supply and bridge unit, right

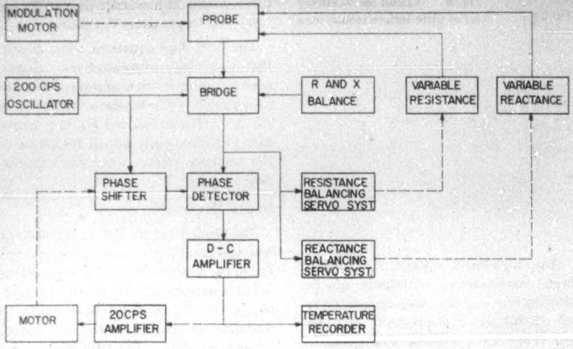


Fig. 10. Block diagram of temperature measuring instrument

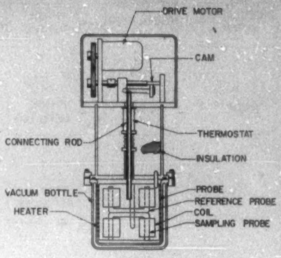


Fig. 11. Inductive probe details

Typical Values

Relative Amplitudes of E_{oT} and E_{oQ}. In the instrument

$d \cong 0.5$ inch
$100 \text{ C} < T < 500 \text{ C}$

in presence of molds of stainless-steel type *420*

R initial = 700 ohms
ωL initial = 300 ohms
$E_{in} = 5$ volts
$E_{oT}/\delta T = 6 \ 10^{-6}$ volt per degree centigrade
$E_{oQ}/\delta d = 2 \ 10^{-3}$ volt per 0.01 inch

Effect of Nonquadrature of E_R and E_{oQ} If the phase difference is $\pi/2 - \epsilon$ instead of $\pi/2$, an error voltage of the order of ϵE_{oQ} is added to E_{oT}. For example:

$$\delta d = 0.01 \text{ inch}; \ \frac{E_{oQ}}{E_{oT}} = 333$$

if $\epsilon = 10$ minutes; $\epsilon E_{oQ} \cong 1$ C

If the distance fluctuations are plus or minus 0.01 inch around the initial position, and if the temperature error is to be smaller than 1 C, the servo system has to adjust the phase angle of the reference voltage with an accuracy of the order of 10 minutes.

Discussion

Correct operation of the instrument is obtained only if the constant temperature curves in Fig. 5 are straight lines within the operating range. This condition is fulfilled by proper design of the sensing inductor. It will be noted that it is not necessary for the constant distance lines to be straight within the operating zone.

Description of Instrument

Equipment Details

The industrial version of the instrument is shown in Fig. 9. The equipment is specially adapted for mold temperature surface measurements for temperatures up to 500 C. The probe can be lowered into the measuring position, in the vicinity of the surface, by means of an air cylinder. When in the upper position, the impedance of the probe is not affected by the presence of the molds.

The block diagram, Fig. 10, schematizes the over-all configuration of the circuit which is derived from the basic block diagram, Fig. 8.

The following features have been incorporated to obtain a high degree of stability in the instrument readings. The impedance of the sensing inductor is periodically compared to that of the reference inductor whenever the probe is in the up position. Any impedance difference caused, for example, by aging or to a variable temperature gradient, is compensated for by the resistance and reactance components.

When the probe is in the temperature measuring position, additional balancing components (R and X balance as shown in Fig. 10) are inserted in the arms of the bridge in order to balance it for the initial or reference temperature and distance (point D in Figs. 4 and 5). These two components are adjusted once in a given process and are mounted in a constant temperature oven.

The probe is represented in Fig. 11. The reference and sensing inductors are placed back to back, embedded in an epoxy resin. The embedment is wound with a heating winding and the entire assembly is suspended in a vacuum bottle. The probe temperature is stabilized at 50 C. The probe is vibrated axially with a maximum amplitude of the order of 0.001 inch by a constant-speed cam and a motor mounted in the upper part.

Instrument Performances

The instrument can operate continuously and detect surface temperature variations of the order of 1 C at temperatures up to 500 C. The over-all time of response is about 5 seconds; although, when balanced, a 5 C per second rate of change of mold surface temperature can be followed. The output is insensitive to spacing variations of 0.01 inch. The readings are emissivity free and the measurements are performed without contacting the mold surface.

Other Applications

The measurement technique is applicable to temperature measurement of conductive nonmagnetic bodies. In this case, if surface temperature measurements were desired, it would be necessary to operate at higher measurement frequencies. The method is equally applicable to electrodeless conductivity measurements.

References

1. Induxtive Temperaturmessung, G. Keinath. *Archiv fur Technisches Messen und industrielle Messtechnik*, Munich, Germany, Jan. 1934, pp. 215–22.

2. Ferromagnetism (book), R. M. Bozorth. D. Van Nostrand Company, Inc., Princeton, N. J., 1951, pp. 769–71.

◆

Discussion

Donald L. Waidelich (University of Missouri, Columbia, Mo.): The method of measuring temperature by the use of an eddy-current probe coil described in this paper by Mr. Mouly is very interesting. The application of a phase detector to minimize the effect of the variation of the spacing of the probe from the metal surface is very similar to the method employed earlier by Dr. F. Förster.[1] In modulating the spacing of the probe from the metal, Mr. Mouly has introduced a useful way of providing a voltage so that the phase relationships can be adjusted readily. Another method[2] of neutralizing the distance component of the voltage utilizes a compensating circuit with the output obtained from an electronic phase meter.

REFERENCES

1. THEORETICAL AND EXPERIMENTAL BASIS OF THE ELECTROMAGNETIC SORTING ACCORDING TO QUALITY OF STEEL SEMIFINISHED PRODUCTS AND STEEL PARTS, F. Förster. *Zeitschrift für Metallkunde*, Stuttgart, Germany, vol. 45, Apr. 1954, p. 206.
2. MINIMIZING THE EFFECT OF PROBE-TO-METAL SPACING IN EDDY CURRENT TESTING, C. J. Renken, Jr., D. L. Waidelich. *Special Technical Publication no. 223*, American Society of Testing Materials, Philadelphia, Pa., 1958, p. 181.

R. J. Mouly: The application of a phase detector to minimize the effect of the variation of the spacing of the probe from the metal surface was employed by G. Keinath and was mentioned in the literature as early as 1934.[1]

Simple-phase detectors of the type used by G. Keinath or F. Förster, however, do not offer a satisfactory solution to the temperature measurement problem described in the paper. They provide an output insensitive to distance variation within a narrow temperature range; theoretically the insensitivity would be achieved for one temperature only. The continuous, automatic adjustment of the phase of the reference voltage becomes a necessity if measurements are to be made over a wide temperature range. This is the essential feature of the instrument.

Naturally, these considerations are applicable to the inductive measurement of electric or magnetic properties whenever it is desired to perform reproducible measurements within a wide range of values.

REFERENCE

1. See reference 1 of the paper.

The Inductronic Electrodynamometer for the Precise Measurement of Voltage, Current, Power, and Energy

R. F. ESTOPPEY
MEMBER AIEE

THE Inductronic* electrodynamometer has been perfected and retested in its final form, and complete information concerning its design and performance can be given at this time. An interim report[1] has been presented covering much of this material, and this paper includes the prior material as well as additional new information. A complete wiring diagram is now included along with output response-time curves and a means of controlling the damping by an external network. Shielding from the effects of external magnetic fields has been improved along with more precise frequency compensation. The network used to obtain this compensation is included.

General Description

The Inductronic electrodynamometer was developed for use as a product resolver[2] for the multiplication of two electric inputs to produce an output current accurately related to the product of the two input currents. The capabilities of this unit have now been extended to include the measurement of voltage, current, and power at 0.1% rated accuracy on direct and alternating currents to 2,500 cps (cycles per second). Transfer error between direct and alternating current is less than 0.04% with readability and resolution of better than 0.02%. Response is fast and power loss is less than that of a conventional dynamometer.

These features have been accomplished by combining an electrodynamometer and a permanent-magnet moving-coil mechanism with the Inductronic amplifier.[3,4] The combining of the two mechanisms is similar to the Shotter-Hawks Comparator,[5] but the addition of the amplifier provides direct current which automatically provides a balancing torque, whereas the comparator requires the introduction of a known direct current for balance.

The product resolver consists of a dynamometer mechanism and a permanent-magnet mechanism having their moving coils mounted on a common staff. The torque of the dynamometer element is automatically balanced by a current fed back into the d-c moving coil from a d-c amplifier, arranged to provide a full feedback current. This amplifier supplies the current needed to produce the counter torque which keeps the system in balance.

Referring to the block diagram in Fig. 1, the dynamometer element is shown at the top with the d-c or torque balance induction element directly below.

The d-c mechanism is supplied with powdered iron inserts set into solid iron pole pieces and core. The magnet produces a steady flux linking the moving coil by means of the solid iron structure, and a 200-kc field is produced in the powdered iron inserts by means of the induction field coils which are attached to the magnet structure. The excitation for the induction field is supplied by an oscillator included within the amplifier, which also amplifies any a-c signal, about 200 kc, induced in the moving coil. With no current in the dynamometer coils, the moving element is kept in a position where the d-c moving coil does not link the 200-kc flux. As current is applied to the dynamometer coils, a torque is produced causing the moving element to turn. The d-c moving coil then links the 200-kc flux and produces a signal which is amplified, rectified, measured, and fed back as a

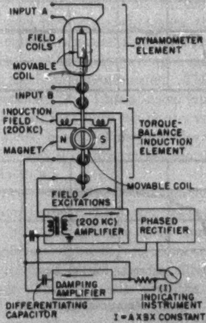


Fig. 1. Schematic diagram of the Inductronic electrodynamometer

---

* Registered trade-mark.

Paper 59-165, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 5, 1958; made available for printing December 23, 1958.

R. F. ESTOPPEY is with Weston Instruments, Division of Daystrom Inc., Newark, N. J.

direct current to produce a torque exactly equal to the dynamometer element torque.

The torque produced by the dynamometer element is directly proportional to the in-phase product of the input currents to the dynamometer coils. An equal counter torque is developed which is proportional to the d-c output current. Therefore, the in-phase product of the input currents is proportional to the output direct current. This direct current varies from 0–1 ma (milliampere) and can be measured by a d-c indicating instrument, precision potentiometer, digital voltmeter, recorder, etc. It can also be used for control purposes.

Damping is supplied by a derivative damping amplifier since the damping of the d-c mechanism is insufficient because of the relatively high moment of inertia of the moving element and high electric torque applied to the moving system. The amplifier responds to the output current change and feeds back a rate-of-change counter current to the d-c moving coil to effectively oppose the change.

Fig. 2 shows the change in output resulting from a sudden application of a constant input to the dynamometer section. Curve *A* is the normal output that can be expected when the output circuit is a resistor. Curve *B* is the voltage drop, taken across an output resistor of 5,000 ohms shunted by a 5-μf (microfarad) capacitor. Curve *C* is the same as *B*, except for the capacitor which is 10 μf. These curves show that the response to a change is very rapid and the damping factor measured on the first overswing is approximately 3. Response time is approximately 0.3 second under normal conditions and can be reduced to 0.2 second by an external *R-C* network. The damping factor is also increased by the external network. The response time is considered more than adequate, since it is faster than the d-c instruments used to measure the output current and many times faster than the present dynamometer instruments having long pointers.

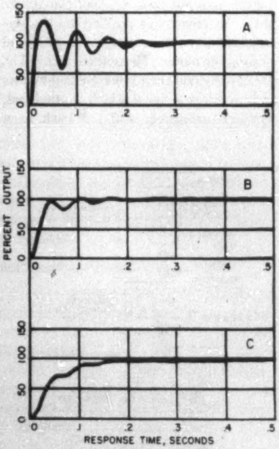


Fig. 2. Response time curves

A—Standard conditions
B—5,000-ohm 5-μf R-C network
C—5,000-ohm 10-μf R-C network

The product resolver unit with cover removed is shown in Fig. 3. The dynamometer element is not visible since it is surrounded by the ferrite shield held in place by the two upper straps. Below the shield can be seen the 200-kc field coils attached to the front section of the magnet structure. Two leveling screws and level are provided to properly level the unit.

The complete wiring diagram is shown in Fig. 4. The resolver unit is shown within the dotted lines. To the right of the unit is the amplifier, oscillator, and frequency discriminator with rectifier. To the left of the unit is the damping amplifier.

Briefly, the operation is as follows: The a-c voltage induced in moving coil (C_1-C_2) is impressed upon the primary of *T-101* where it is stepped up and is applied to the grid of the amplifier tube *V-101*. Here it is amplified in two stages.

The oscillator consists of *V-102* and the second stage of *V-101*. The tank circuit consists of the 200-kc field coils (F_1-F_2) in series with transformer *T-103*, both of which are tuned to cause the oscillator to oscillate at 200 kc with no input from the preamplifier. Regenerative feedback for the oscillator is by means of the *510k* resistor and the second stage of *V-101*.

The amplified voltage is introduced into the oscillator circuit. This voltage is essentially in quadrature with the oscillator feedback voltage (*510k*) and causes the oscillator frequency to change in order to maintain proper phase relationship for oscillation. The output voltage from transformer *T-103* feeds into a frequency discriminator as commonly used as a detector in frequency-modulated radio receivers. This rectifies the a-c voltage and produces a direct current which is a function of the oscillator frequency. This direct current is then fed back to the moving coil to develop torque to balance the system.

Damping is provided by an *R-C* network and a single-stage amplifier *V-701*. The amplifier responds to the d-c output current change by amplifying the voltage obtained from the *1k* resistor in series with the moving coil. The output of the damping amplifier provides a rate-of-change current which is fed back to oppose the input change of the d-c moving coil. The time constant of the circuit is controlled by the 0.1-μf capacitor and the circuit resistance.

Accuracy and Frequency Errors

Accuracy is determined by the design and stability of the dynamometer and d-c mechanisms. The stability of the amplifier is not a factor because the very high gain of the system and the large feedback ratio. This is a very important part of the design since the amplifier must not be a source of error. Its function is to supply and feed back sufficient current to balance the system. Aging of tubes or other changes which may change the amplifier gain merely cause the moving coils to deflect more than it did previously so as to maintain the same output current which is fed back to keep the system in balance.

The moving element consists of two round moving coils with a through shaft to minimize twisting as torque is applied to the coils. Four conducting filaments and mechanical balancing means complete the moving element to form a stable design. Since mechanical spring torque is not used, spring fatigue is not a factor. Stability has been obtained by conventional high quality design of the d-c and dynamometer elements.

Rated accuracy is 0.1% of full-rated d-c output. Ready means of adjustment to better than 0.1% is obtained by an ad-

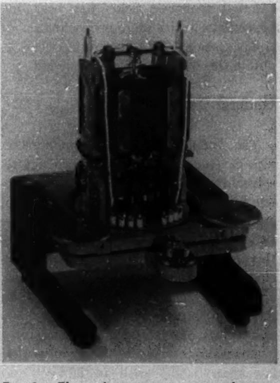


Fig. 3. Electrodynamometer unit with cover removed

Vol.
Pt. 1
1959

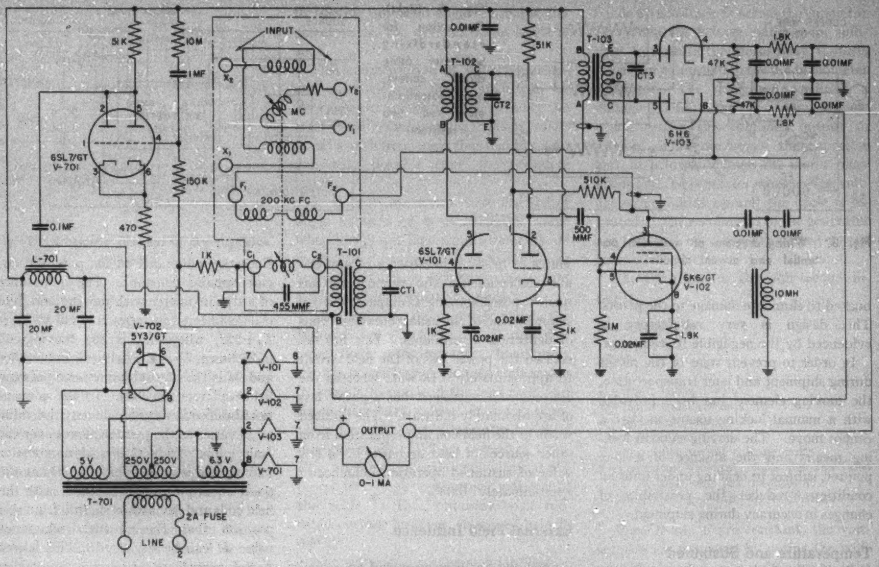


Fig. 4. Complete wiring diagram of Inductronic electrodynamometer

justing screw which changes the position of a magnetic shunt on the d-c element magnet. Direct-current reversal errors are made negligible by a balance adjustment which compensates for residual magnetism in the surrounding magnetic parts.

The output current is directly proportional to the in-phase product of the input currents over a wide range. Calibration errors, as found in an instrument with a graduated scale, do not exist. For example, correction factors obtained from testing a wattmeter over a span of 100–500 watts showed a difference of less than 0.03% of reading.

The very slight movement of the moving element also makes it possible to keep the angle between the moving and field coils of the dynamometer element approximately 90 degrees apart at all times. The field-coil flux will thus cause little or no voltage to be induced in the moving coil. Since this induced voltage increases with frequency, it must be kept small in order to reduce the frequency error to a negligible value. This source of frequency error is called the mutual inductance error and has been effectively eliminated in the dynamometer by locating the moving coil at its point of zero mutual as described previously.

Another source of frequency error is caused by eddy currents circulating in the metal parts which are linked by the field-coil flux. This shows up as a lower indication as the frequency is increased since the eddy currents increase with frequency. It also causes the field-coil flux to lag the current producing it. The angular difference between the field-coil flux and current is called the current-circuit defect angle. This will cause an error in wattmeter indications at the lower power factors and must be kept small.

The eddy currents in the product resolver have been kept very small by surrounding the field coils with a ferrite shield and by using insulated strands of wire on field coils of few turns where a large cross section of wire is required. Most of the field coil flux is confined within the shield which has a very high resistance to eddy currents. As a result of this construction, the eddy current error has been kept very small.

A typical wattmeter was tested at the National Bureau of Standards to determine the difference between indications on alternating current and reversed direct current. The difference was undetectable at 60 cps, 0.03% high at 400 cps, 0.01% high at 800 cps, and 0.04% low at 1,600 cps. The current-circuit defect angle has been measured and is approximately 7 minutes at 2,500 cps, decreasing to a negligible value at 60 cps.

Other frequency errors are caused by the inductance of the dynamometer coils. This has been compensated and is described later for each type of measurement, i.e. voltmeter, wattmeter.

The accuracy of indication is also dependent upon the accuracy and readability of the output instrument. For the most precise measurements, the output current is passed through a precision resistor and the voltage drop is measured by a precision potentiometer. The direct current output is 1 ma at rated input and the output resistance can be any value up to 5,000 ohms. This provides ample power for accurate measurement and readability.

Friction errors are approximately 0.01% of rated output. This error has been kept low by maintaining a good torque-to-weight ratio of the moving element. It is also reduced further by the alternating torque always present from the 200-kc field. Because of these factors, it has been possible to use a pivot and jewel bearing design rather than a suspension design. Pivots are of high-quality steel and the jewels are of sapphire, spring-

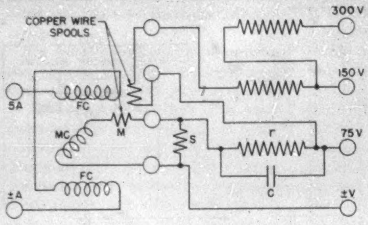


Fig. 5. Wiring diagram of wattmeter potential and current circuits

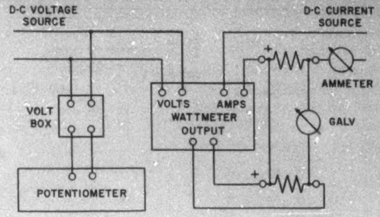


Fig. 6 (right). Connection for standardizing wattmeter using direct current, one potentiometer and two resistors

backed to eliminate damage to the pivots. This design is very satisfactory as evidenced by the negligible friction error.

In order to prevent wear of the pivots during shipment and later transportation, the moving element has been provided with a manual locking means so that it cannot move. The moving element locking means and the absence of a long pointer, subject to bending under unusual conditions, reduces the possibility of changes in accuracy during shipment.

Temperature and Sustained Operation Influence

Changes in temperature will change the resistance of copper-wound coils in the dynamometer mechanism and will change the flux in the magnet of the d-c mechanism. These will then cause errors in indication as the ambient temperature changes or as the coil resistance changes as a result of self-heating and heating of adjacent parts such as the series resistor. The effect of the changes in magnet flux with temperature is to increase the output approximately 0.016% for an increase of one centigrade degree in temperature. Copper resistance in the potential circuit of the voltmeter and wattmeter will cause the output to decrease for an increase in temperature. By proper selection of the ratio of copper resistance to total series resistance, it has been possible to make the potential circuit change approximately equal to the magnet-flux change and thus compensate for changes in ambient temperatures.

The sustained operation influence, or effect of self-heating, is also reduced to a very small amount as a result of the temperature-compensated design and also by the absence of spring fatigue since springs are not used. In addition, the product resolver unit is in a separate closed container completely removed from the series resistance and amplifier, both sources of heat. This, then, leaves the field coils in the product resolver unit as the main source of heat. The absence of appreciable deflection for the moving coil has made it possible to use a compact field coil shape instead of the usual round field coils of deflection instruments. This has decreased the power loss of the field circuit to approximately 0.18 watt whereas the conventional wattmeter has a power loss of approximately 0.8 watt. The reduced watts in the field coil and separation from other sources of heat accounts for a low value of sustained operation influence of approximately 0.03%.

External Field Influence

Shielding has been provided by a steel case around both mechanisms. In addition, a single wrap of 0.018-inch silicon steel has been added around the dynamometer mechanism. This has provided very effective shielding from external magnetic fields and the effect of a 5-oersted field is less than 0.1% of rated output when used on alternating current.

Voltmeter

The voltmeter uses moving and field coils rated at 50 ma for a multirange instrument of 300/150/75 volts. Series resistance consists of mica sheets wound with low-temperature coefficient resistance wire in a manner to be practically noninductive and capable of maximum heat dissipation. The d-c output is proportional to the output voltage squared.

Where required, frequency compensation is added by shunting the low-range series resistance with a capacitor. This method of compensation has been discussed in a prior paper.[6] Compensation is practically flat to 2,500 cps because of the low ratio of reactance to resistance in the potential circuit. Calculations for the 75-volt range show $X_L/R=0.26$ and $r/R=0.95$ at 2,500 cycles. From the curves of Fig. 4 of the referenced paper, the maximum frequency error is determined as less than 0.01%.

This compensation is also very exact for all outputs of the Inductronic voltmeter since the inductance in the potential circuit remains constant. The inductance of a dynamometer with moving and field coils connected in series is equal to L_m+L_f+2M, where L_m is the moving-coil inductance, L_f is the field-coil inductance and M is the mutual inductance between field and moving coils. In the conventional deflection dynamometer, the value of M varies as the pointer moves over the scale. The Inductronic dynamometer moving coil is at or approximately at the point of zero mutual inductance with the field coils and deviates so slightly from this position that the mutual inductance value M can be neglected. This leaves $L_m+L_f=$constant and hence, no change in potential circuit inductance. This means there is no scale error and the previously mentioned compensation for frequency at one scale point, or output in the case of the Inductronic voltmeter, will provide the same compensation at other scale points.

As a result of a low value of constant inductance, compensation can be very exact and the Inductronic voltmeter offers better than 0.1% accuracy from direct current to at least 2,500 cps.

Ammeter

Low current ranges of 15–150 milliamperes are provided by winding the moving and field coils with the appropriate turns and connecting them in series. Like the voltmeter, the d-c output varies as the square of the input current. If desired, the moving and field coils can be brought out to separate terminals and the instrument can then be used as a product resolver or current product meter.

Higher current ranges to one ampere are possible by means of a shunt across the moving coil. In order to keep the frequency and temperature errors to a low value, it is necessary to add resistance in series with the moving coil and shunt the combination. This requires approximately 5 volts across the shunt and makes ranges above 1 ampere impractical except by the use of a current transformer.

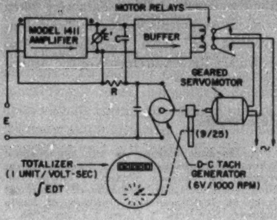


Fig. 7. Schematic diagram of the precision integrator

Wattmeter

The wattmeter uses a moving coil having a nominal rating of 25 ma connected in series with resistor sheets as used in the voltmeter. The field coil is wound for 5 amperes or lower.

In the section headed "Accuracy and Frequency Errors," it was explained that the eddy current errors are kept very small. Another source of error in the wattmeter is the defect angle in the potential circuit. This is caused by the inductance of the moving coil which causes the potential circuit current to lag the applied voltage by an angle equal to $\tan^{-1} L_m/R$ where R is the potential circuit resistance and L_m the moving-coil inductance. This can be compensated by shunting the series resistance with a capacitor computed from the equation $C = L_m/r^2$. This equation is given in reference 6, where r is equal to the resistance shunted by capacitor C. This compensates for the potential-circuit defect angle leaving the current-circuit defect angle caused by the eddy currents.

This angle can be compensated by shunting the moving coil with a very high resistance as shown in Fig. 5.

Let

I = current coming into shunted moving coil circuit
i = moving coil current
M = total moving coil resistance which is shunted
S = shunt resistance
X = moving coil reactance at the compensated frequency
θ_s = phase angle between I and i

$$i = \frac{IS}{S+M+jX} = \frac{IS(S+M)-jISX}{(S+M)^2+X^2} \qquad (1)$$

$$\tan\theta_s = \frac{j \text{ terms}}{\text{real terms}} = \frac{-X}{S+M} \qquad (2)$$

This is then the angle by which the current in the moving coil lags the applied current in the potential circuit. The current-circuit defect angle is also lagging and by making the shunt angle θ_s equal to the defect angle, the over-all defect angle is zero.

This compensation of both potential- and current-circuit defect angle can be very exact to as high as 2,500 cps since the error and compensating means vary directly with frequency, thus making compensation possible over the complete range.

It is very important that the defect angle be reduced to a negligible value when low power-factor measurements are made. For example, an over-all defect angle of 2 minutes will cause an error at 10% power factor of 0.6% of reading.

The negligible defect angle makes the Inductronic wattmeter ideal for low power-factor measurements. In addition, the low power loss in the field coils makes it possible to overload the field coils four times their rated value. It is also possible to use a d-c output indicator having a full-scale range of 0.2 ma instead of 1 ma. This provides full-scale readability at 5% power factor. The accuracy will not be the same as that obtained with full-rated output, however, it can be held to 0.5%.

The wattmeter can be standardized on direct current by using a precise voltage source and two four-terminal resistors or shunts, see Fig. 6. One shunt is connected in series with the current circuit and the other is connected across the d-c output terminals. The two shunts are connected together so as to compare voltage drops by means of a sensitive galvanometer. Direct current of the approximate value desired is applied to the field coils and a direct voltage is applied to the potential circuit. This voltage is varied until the galvanometer between the shunts reads zero. The error is then the difference between the voltage applied and the rated voltage. Changes in current will not affect the balance since the voltage drop on both resistors will go up and down together for changes in current. By this means, the wattmeter can be standardized by using one precision potentiometer.

Watt-Hours

Watt-hours are measured by connecting the output current of the wattmeter to a resistor to provide a millivolt drop of 0–50 millivolts. This voltage is then applied to a precision integrator which measures the time integral of the voltage.

Such an integrator is shown in Fig. 7. Input voltage E is applied to the amplifier which produces an output current proportional to the input millivolts. This builds up a voltage E' across the capacitor C. When the voltage build-up is sufficient to operate the motor relay, it starts the motor connected to the d-c tachometer generator. The generator produces a voltage greater than the input voltage which causes the amplified output to reverse. This discharges the capacitor and, at a lower voltage the motor relay opens and the generator stops. The voltage again builds up and the cycle is repeated. The revolutions of the generator are totalized on a counter geared to the generator.

The equation for the volt-second input is

$$Edt = RCE' + k\theta$$

where

E = input voltage, volts
dt = input time, seconds
R = resistance, ohms
C = capacitance, farads
E' = amplifier output voltage, volts
k = generator constant, volts per rps (also volt-seconds per revolution)
θ = rotation of generator in revolutions

Since R and C are constant, the volt-seconds are measured by noting the voltage E' recorded on a voltmeter and the generator revolutions given on the counter. By suitable calibration and selection of constants, the meter and counter are calibrated in terms of volt-seconds, watt-hours, etc.

For long periods of time, such as 10 minutes or more, the volt-seconds stored in the capacitor at the end of a cycle do not have be read on the voltmeter. This

Table I. Summary of Wattmeter Characteristics

Requirement	Inductronic Wattmeter	ASA C39.1
Rated accuracy, % of full scale	0.1	0.1
Damping factor	3	2.5 min
Response time, seconds	0.3	10 max
Loss, volt amperes:		
150-volt range at 120 volts, 60 cps	3.6	5 max
5 ampere range at 5 ampere, 60 cps	0.2	2 max
External temperature influence % for ±10 C		
75 volts	negligible	0.1 max
150 volts	negligible	0.1 max
300 volts	0.08	0.1 max
Sustained operation influence, %	0.03	0.1 max
External field influence, %, 5-oersted field	0.1	0.75 max
Frequency influence, d-c to 66 cps, %	less than 0.01	0.05* max
Voltage influence ±10% variation, %	less than 0.01	0.1 max
Power factor influence, 1.0–0.5 pf, 60 cps, %	negligible	0.1 max

* The error allowed is for a lesser change to 54–66 cps from 60 cps.

is because the counter indication is large and the small amount of volt-seconds indicated on the voltmeter can be neglected. For shorter periods of time, the sum of the voltmeter and counter readings should be added together to give the total volt-seconds.

The basic accuracy for normal periods of time (10 minutes or more) is completely dependent upon the stability of the generator. This stability has been obtained by good design practice and 0.1% accuracy is maintained. For short periods of time, the accuracy of the instrument and readability of the counter are a factor and must be considered. A more complete description of the precision integrator is given in reference 7.

The combination of Inductronic wattmeter and integrator has been tested under actual field conditions to determine its suitability for measuring turbine efficiencies. This test was performed by the engineers of the Long Island Lighting Company at their Barrett Power Station.

Five tests were run with indicating wattmeters in each of three phases and with one Inductronic wattmeter and integrator in phase A. The test time varied from 1–3 hours. The integrator was read at the beginning and end of each test and the wattmeters were read every minute.

The largest difference between the two watt-hour readings by each method was 0.12%. The average difference for all five tests was 0.08%. It is believed that using Inductronic wattmeters in the other two phases and feeding the output to three integrators, would have provided total generator output in kilowatt-hours to a accuracy of approximately 0.1%. The output of the three wattmeters can also be connected to one integrator to provide the total 3-phase energy. This may not be convenient, where transformer correction factors must be applied for each phase and where it is desirable to know the load balance.

Conclusions

The product resolver can be used for the measurement of volts, amperes, watts, and watt-hours when used in conjunction with a precision integrator. It offers 0.1% accuracy over a wide power-frequency range and provides a d-c output which can be measured or used for control. Summarized in Table I are some of the important characteristics of the Inductronic wattmeter. These data were taken from measurements made on a typical instrument or from calculation. Also shown are the requirements of the American Standards.[8]

References

1. An Interim Report on the Inductronic Electrodynamometer for the Precise Measurement of Voltage, Current, Power and Energy, **R. F. Estoppey.** *Transactions*, Professional Group on Instrumentation, Institute of Radio Engineers, New York, N. Y., vol. 1–7, Dec. 1958.

2. Weston Inductronic Product Resolving System, **R. W. Gilbert.** *Weston Engineering Notes*, Weston Electrical Instrument Corporation, Newark, N. J., vol. 12, no. 3, Sept. 1957, pp. 4–6.

3. The Induction Galvanometer, a Sensitive Instrument Converter, **R. W. Gilbert.** *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1121–26.

4. The Inductronic Servo Amplifier, **R. C. Langford.** *Weston Engineering Notes*, Weston Electrical Instrument Corporation, vol. 12, no. 3, Sept. 1957, pp. 1–6.

5. A Precision AC/DC Comparator for Power and Voltage Measurements, **G. F. Shotter, H. D. Hawks.** *Journal*, Institution of Electrical Engineers, London, England, vol. 93, pt. 2, no. 34, Aug. 1956, pp. 314–24.

6. Frequency Compensation of A-C Instruments, **John H. Miller.** *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 217–21.

7. The Weston Model 1473 Precision Integrator, **R. W. Gilbert.** *Weston Engineering Notes*, Weston Electrical Instrument Corporation, vol. 10, no. 2, Dec. 1955, pp. 6–8.

8. Electric Indicating Instruments. *ASA C39.1-1955*, American Standards Association, New York, N. Y., Aug. 1955.

Vol.
Pt. 1
1959

Polyphase Meter Connections 30 Years After Woodson

F. W. WARBURTON
MEMBER AIEE

Synopsis: In 1928, W. C. Woodson announced his procedure for verifying instrument transformer connections to meters for 3-phase 3-wire services.[1] The Woodson method has earned the reputation of being the most reliable and satisfactory practice for proving these connections. In 1928, the procedure had minor limitations. Since that time, the employment of 3-phase 4-wire distribution systems, the growing impracticability of reducing customer loads for connection testing to balanced 3-phase known impedances, and the increasing use of capacitors have caused difficulty both in proving internal meter connections and in estimating the load phasors needed for comparison with the Woodson phasors obtained at the meter test switch. There is no procedure that will signal a definite "no" or "yes" that meter test switch phasors correspond to load phasors. On the contrary, the matching of load and Woodson phasors is entrusted to human judgment, and errors are probable. On one utility system, 3$^1/_2$ stator Woodson kits, used in proving over 1,500 complex metering installations, have given added but not complete insurance for verifying 4-wire Y systems. Thirty years of electrical growth and changes have increased the limitations to proving transformer-connected meter installations. In the meantime, there have been no adequate new or extended procedures to give required integrity in this area of metering.

THE METERING of large customers of electric power utilities involves the use of potential and current transformers with secondary connections to the meter. Depending upon individual company practices, and whether 3-wire or 4-wire service is provided, usually five to eight current and potential-transformer secondary conductors are connected to the meter test switch. In making connections with these transformer secondaries, several errors are possible; wrong transformer ratio or polarity, transformers connected in or to the wrong primary conductors, transformer secondary connections reversed, grounded, opened, or short-circuited. The possibility of error is increased by making installations under emergency conditions, the use of nonstandard materials, incorrect reconnection of leads opened during a test, and the absence of key personnel. Another possible contribution to error is the fact that most metermen consider that each stator of a 3-stator meter on a 4-wire Y system always should cause the disk to run in a "forward" direction.

Electric power companies, soon after

Paper 59-168, recommended by the AIEE Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted August 14, 1958; made available for printing December 2, 1958.

F. W. Warburton is with the New England Power Service Company, Worcester, Mass.

The author appreciates the contributions, in both data and ideas, of his colleagues G. A. Palmer, L. S. Salmonsen, and J. F. Stasaitis of the New England Electric System.

installing or modifying metering facilities to their large energy-taking customers usually have a specialist make a verification check of instrument transformer connections. To uncover any faults in the connections, some companies depend upon visually tracing the wiring and making meter stator tests with use of customer equipment for a test load. For this purpose, wire coverings can be colored or have printed number identification. The new plastic-sheathed conductors offer good legibility, with promise of more permanence than has formerly been obtained. However, because of pole-top-mounted transformers, concealed wiring, wrong nomenclatures, human errors in tracing wires, and hazards to employees, other companies have substituted special checks, such as the Woodson method applied at the meter test switch, as the best means to insure integrity of connections.

Since internal meter connections also can be improper, the connection verification checker in many companies proves meter-coil integrity as well as the wiring to the instrument transformers. For complete audit, a meter calibration test, register proving, and a register-multiplying constant check are required. This study attempts to show that current practices used for verification of connections are not precise operations and that the laws of probability point to serious, if infrequent, errors.

The ultimate purpose of the Woodson procedure or other connection verification is to insure correct billing to customers. If a utility should install a meter with an improper instrument transformer connection and subsequently "approve" it by a Woodson or other test, the error in billing could continue indefinitely. This situation results because generally there are no previous bills suitable for comparison and most line loss studies, if made, must be more nearly exact than is normally practical to uncover such an error. As a consequence, the mistake may well continue for the life of the installation, a period that may cover 25 years or more. A discrepancy exceeding several million kilowatt-hours a year may fall in this category.

Since a watt-hour meter measuring energy at the usual balance of load and at a power factor of 0.7 lag to 1.0, normally causes highest registration when connections are correct, it is not surprising to find that nearly all connection errors penalize the supplier of energy. On one system, eight connection errors uncovered over a period of 5 years were all causing low measurements; when they were corrected, an estimated $60,000 was added annually to energy revenue.

For brevity, the discussion except where noted will assume that: metering follows Blondel's theorem; no instrument transformer secondary conductor is employed for both "potential" and "current" use; meter test switches are inserted immediately before the meter and at the terminations of the instrument-transformer secondary conductors; only the original Woodson procedure for instrument-transformer connections and the more popular meter stator test procedures are employed; the common and grounded point of the current-transformer secondaries is or can be made available at the meter test switch.

The Woodson Method

OBJECT

The Woodson contribution to metering is a procedure whereby opportunity is presented for a tester and any reviewing engineer to decide whether the phasor diagram constructed from currents and voltages measured at the meter test switch agrees or does not agree with the estimated load phasors. If the Woodson and load phasors are in agreement, the connections from the primary supply will provide the metering quantities, in proper magnitude and order, to the meter test switch. However, as yet no mechanical procedure proving the correctness of the diagram construction can supply proof that the diagram agrees or disagrees with the load phasors. The question of agreement must be answered by man. As already stated, the serious present weakness of the system is the occasional difficulty in estimating load phasors with required accuracy.

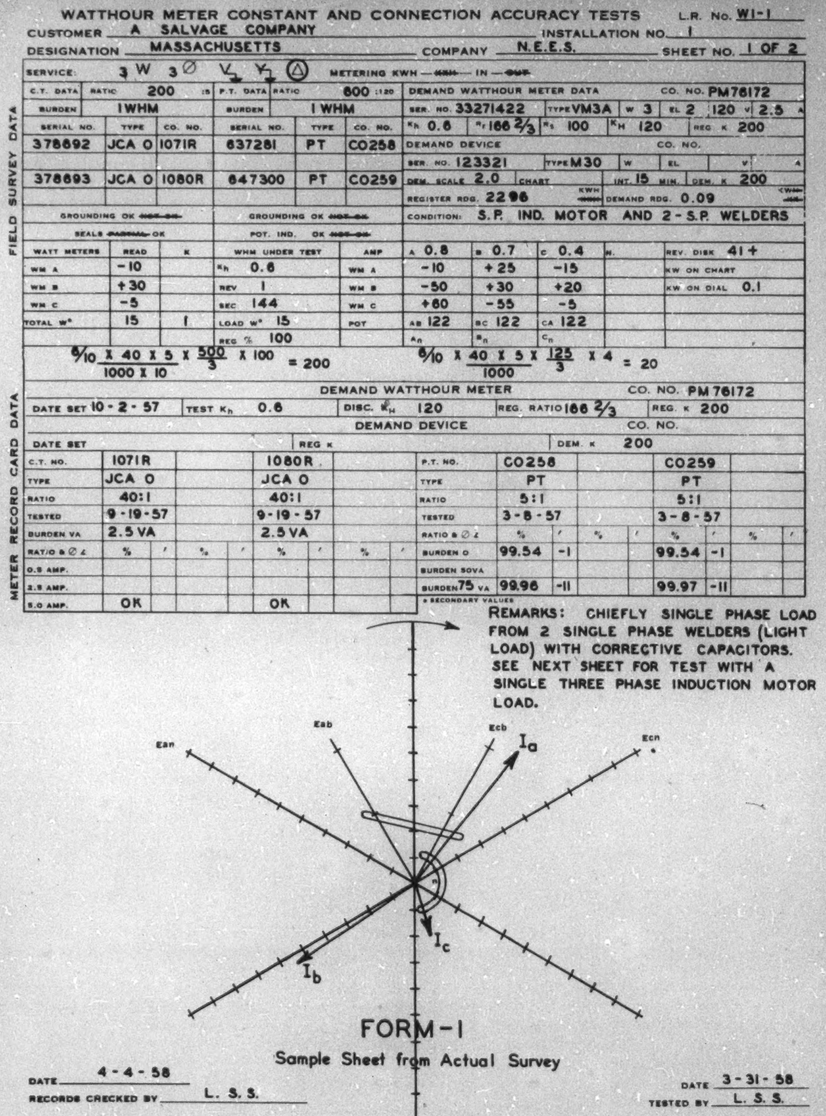
SCOPE OF TEST PROCEDURE

Woodson necessarily used contemporary meter test procedures to supplement his check of transformer connections and he thereby determined the soundness of the complete metering installation. Thus Woodson's procedure concerned transformer connections only, but to provide an acceptable degree of assurance that energy and power registrations are satisfactory, it is also necessary to trace connections between meter switch and meter, verify internal meter connections, check meter registration, determine instrument-transformer performance, and ascertain correct multiplying constants. In the author's company, watt-hour meter constant and conection accuracy tests comprise: meter current coil continuity, single stator, and Ackerman tests; Woodson "star connection" wattmeter checks against meter disk revolutions; establishment of the phasor diagram of currents and potentials at the meter test switch; comparison of this diagram with the estimated load; decision as to correct, incorrect, or inconclusive connections; calculation of multiplying constants; inspection of instrument transformers to ascertain proper capacity and relative location of conductor connections; inspection for safety, approved devices, utility of layout, and possible diversion of energy. These are all accomplished at the meter installation. Later, instrument-transformer test record cards are screened at the operating company office.

PERTINENT PARTS OF TEST PROCEDURE

The installation from meter switch through the meter is first checked. The initial test item is an "as found" visual inspection of meter type, disk speed, and disk direction of rotation. Watt-hour meter dial reading, maximum-demand deflection, and meter nameplate data are recorded on the Form I work sheet in Fig. 1. The second step is a tracing of connections between test switch and meter. Here, the most trouble is encountered when switch connections are not made as specified in the standards. Occasionally, poor switch action will cause difficulty.

Next, partial integrity of the meter is investigated. Action depends upon the type of meter, the load, and the preferences of the tester. The object is to test for coil opens, grounds, reversals, interchanges, and disk constant. Here a 2-stator, *A* and *C* meter with 3-phase 3-wire delta service is assumed. The first step is a combined proving of meter coil continuity and a stator check. A suitable meter test switch would have potential blades *A*, *B* (common), and *C;* current blade pairs *A* and *C;* or a total of seven subswitches. All meter switch blades except potentials *A* and *B* are opened. Woodson current plug *A* is inserted in current series jack *A*, short-circuit current blade *A* is closed, and the Woodson ammeter observed. Also, for economy of motion, current plug *C* is inserted in current series jack *C* at this time, but short-circuit switch blade *C* is not closed. All three potential leads *A*, *B*, and *C* are clipped onto the line side of the respective meter potential switches *A*, *B*, and *C*. Meter disk rotation and wattmeter and ammeter usually indicate continuity of circuitry and "proper" meter-stator *A* performance. Similarly, continuity of


Ean
Eab
Ecb
Ia
Ecn
Ic
Ib
Ebn

WATTHOUR METER CONSTANT AND CONNECTION ACCURACY TESTS L.R. No. WI-1

CUSTOMER A SALVAGE COMPANY INSTALLATION NO. 1

DESIGNATION MASSACHUSETTS COMPANY N.E.E.S. SHEET NO. 1 OF 2

FIELD SURVEY DATA

SERVICE: 3 W 3∅ V Y (△) METERING KWH — ~~KVH~~ — IN — ~~OUT~~

C.T. DATA	RATIO 200 :5		P.T. DATA	RATIO 600 :120	
BURDEN	1WHM		BURDEN	1 WHM	
SERIAL NO.	TYPE	CO. NO.	SERIAL NO.	TYPE	CO. NO.
378692	JCA O	1071R	637281	PT	CO258
378693	JCA O	1080R	647300	PT	CO259

GROUNDING OK ~~NOT OK~~ — GROUNDING OK ~~NOT OK~~

SEALS ~~PARTIAL~~ OK — POT. IND. OK ~~NOT OK~~

DEMAND WATTHOUR METER DATA CO. NO. PM76172

SER. NO. 33271422 TYPE VM3A W 3 EL 2 120 V 2.5 A

Kh 0.6 Rr 166 2/3 Rs 100 KH 120 REG K 200

DEMAND DEVICE CO. NO.

SER. NO. 123321 TYPE M30 W EL V A

DEM. SCALE 2.0 CHART INT. 15 MIN. DEM. K 200

REGISTER RDG. 2296 KWH DEMAND RDG. 0.09 KW

CONDITION: S.P. IND. MOTOR AND 2-S.P. WELDERS

WATT METERS	READ	K	WHM UNDER TEST	AMP	A 0.8	B 0.7	C 0.4	N.	REV. DISK 41+
WM A	-10		Kh 0.6	WM A	-10	+25	-15		KW ON CHART
WM B	+30		REV 1	WM B	-50	+30	+20		KW ON DIAL 0.1
WM C	-5		SEC 144	WM C	+60	-55	-5		
TOTAL W*	15	1	LOAD W* 15	POT	AB 122	BC 122	CA 122		
			REG % 100		An	Bn	Cn		

$\frac{6}{10} \times \frac{40 \times 5}{1000 \times 10} \times \frac{500}{3} \times 100 = 200$ $\frac{6}{10} \times \frac{40 \times 5}{1000} \times \frac{125}{3} \times 4 = 20$

METER RECORD CARD DATA

DEMAND WATTHOUR METER CO. NO. PM 76172

DATE SET 10-2-57 TEST Kh 0.6 DISC. KH 120 REG. RATIO 166 2/3 REG. K 200

DEMAND DEVICE CO. NO.

DATE SET REG K DEM. K 200

C.T. NO.	1071R				1080R			
TYPE	JCA O				JCA O			
RATIO	40:1				40:1			
TESTED	9-19-57				9-19-57			
BURDEN VA	2.5 VA				2.5 VA			
RATIO & ∅ ∠	%	′	%	′	%	′	%	′
0.5 AMP.								
2.5 AMP.								
5.0 AMP.	OK				OK			

P.T. NO.	CO258				CO259			
TYPE	PT				PT			
RATIO	5:1				5:1			
TESTED	3-8-57				3-8-57			
RATIO & ∅ ∠	%	′	%	′	%	′	%	′
BURDEN O	99.54	-1			99.54	-1		
BURDEN 50VA								
BURDEN 75 VA	99.96	-11			99.97	-11		

* SECONDARY VALUES

REMARKS: CHIEFLY SINGLE PHASE LOAD FROM 2 SINGLE PHASE WELDERS (LIGHT LOAD) WITH CORRECTIVE CAPACITORS. SEE NEXT SHEET FOR TEST WITH A SINGLE THREE PHASE INDUCTION MOTOR LOAD.

FORM-1

Sample Sheet from Actual Survey

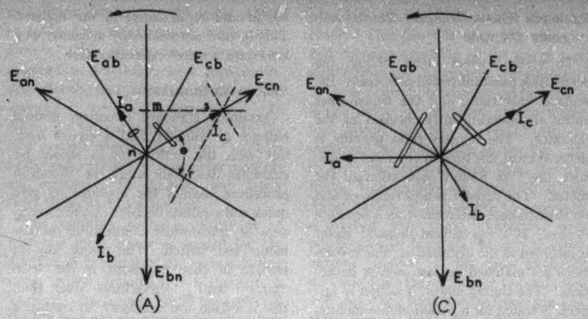
DATE 4-4-58

RECORDS CHECKED BY L. S. S.

DATE 3-31-58

TESTED BY L. S. S.

Fig. 1. Form 1: Sample sheet from actual survey



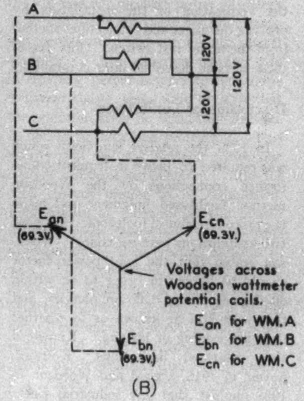


Fig. 2. Example 1: Service, 3-wire delta; type of meter, 3-wire 3-phase 2-stator; Woodson readings 150 watts. Potential voltage for AB, BC, and CA, each 120 volts

(A), (B)—As found:

	A	B	C	Total
Amperes	0.5	1.0	0.866	
WMA	+30		−30	0
WMB	−30	+60	−30	0
WMC		−60	+60	0
Total	0	0	0	0

(C)—As left:

	A	B	C	Total
Amperes	1.0	0.5	0.866	0
WMA	+60	−30	−30	0
WMB		+30	−30	0
WMC	−60		+60	0
Total	0	0	0	0

meter coil *C* and performance of stator *C* are checked. However, if the disk does not turn for either stator *A* or *C* the tester must determine whether a potential coil is open; or one stator is operating at power factor = 0.0; or the load is essentially single-phase and current in *A* or *C* is insufficient to turn the disk. The continuity of the two potential coils can be determined by the Ackerman check. In this test all switch blades except potential *B* are closed and line voltage *AC* is placed across the 2-meter potential coils in series. Currents from transformer-secondary conductors *A* and *C* are supplied to meter coils *A* and *C* respectively. If the disk turns, continuity of both potential coils is assured. Also, if the disk speed increases with later closing of potential switch *B*, added insurance to verification of connections is made.[2] Single-phase loading can be determined from the Woodson ammeter measurements. A power factor of zero per cent will be substantiated by the Woodson diagram.

The meter disk constant test usually follows. The Woodson kit is used as a wattmeter to indicate flow of power to the meter. All meter test switch blades, except the two "jacked" units, are closed with current plugs *A*, *B*, and *C* and potential clips *A*, *B*, and *C* still connected to the switch from the preceding test. For the period of from one to perhaps ten disk revolutions, depending upon load conditions, the Y (star) connection watt meter watt-seconds are calculated, recorded, totalled, and matched against the meter disk constant. If agreement between meter performance and Woodson measurements is not within 5% the run is repeated.

Woodson Phasor Diagram

The phasor diagram data are next obtained as a step in proving connections between the instrument transformers and the meter test switch. To measure these data quickly and easily, Woodson devised his "kit." It usually, for 3-wire services, comprises a voltmeter, three identical wattmeters with multipliers, an ammeter, a phase-sequence indicator, and required switches in a single "black box." The switches reverse polarities and arrange connections in Y or in series. The normal practice is to connect the potential coil leads of the wattmeters to the three potential switches, which are connected in turn to the potential-transformer secondaries. In the 3-wire delta service, this means that the kit wattmeter potential coils are connected in Y to the delta of line potentials *AB*, *BC*, and *CA* as shown in Fig. 2(B). The neutral of the Woodson Y, although not needed in this case, is accessible on the kit for connection purpose. In Fig. 2(A), with a 3-phase delta service, the 3-wattmeter voltages are assumed to be 120 degrees apart, as shown by E_{cn}, E_{bn}, E_{cn}, and are used with kit wattmeters *WMA*, *WMB*, and *WMC* respectively. The assumed secondary load currents I_a, I_b, and I_c are in turn switched through each of the three wattmeters and the single ammeter. Values of current and watts are recorded. This constitutes Woodson's series-connection test.

All service voltages available at the meter test switch are next measured by the kit voltmeter and recorded. If unbalanced or abnormal values are obtained, the investigator is alerted to take additional tests, such as checks for grounds on the potential circuitry or as indicated in the section "Load Voltage Unbalance." Phase sequence is then ascertained and recorded.

In Fig. 2(A), when I_c, with a value of 0.866 ampere, flows through *WMA*, the reading is −30 watts; through *WMB* −30 watts; and through *WMC*, +60 watts. With balanced voltages, E_{an}, E_{bn}, and E_{cn}, the wattmeter readings are proportional to the watt components of the currents I_a, I_b, and I_c. This voltage balance and associated current and watt proportionality permit the phasor diagram to be determined by plotting the wattmeter readings to scale on the voltage phasors. Thus, to determine I_c, −30 watts, representing the watt component of I_c, is plotted to a convenient scale along E_{an}. This actually is +30 watts along $-E_{an}$. At this point a perpendicular *r-s* is constructed. I_c must lie at some point on this line *r-s*. Likewise, with +30 watts plotted on $-E_{bn}$ and +60 watts on E_{cn}, I_c must also lie at some point on these respective $-E_{bn}$ and E_{cn} perpendiculars. It will be noted that all three perpendiculars, *r-s*, *m-s*, and the one erected on E_{cn}, meet at a point. The intersection of the three perpendiculars is

I_c. The currents I_a and I_b are determined similarly. If all three perpendiculars establishing a current phasor do not intersect at a point, either the load has changed during the time the measurements were taken, or some trouble exists in assumptions of voltage phasors or in the mechanics of obtaining and plotting data. Arithmetical checks to nearly zero of the wattmeter tabulations, both horizontally and vertically, are assumed to signify that the load did not change during time of measurement and that the mechanics of making readings and entering table data are correct. The facts that wattmeter tabulations add to zero and that all three perpendiculars do meet at a point are not proof of proper connections, as further explanation will demonstrate.

Since three balanced voltages, E_{an}, E_{bn}, and E_{cn}, are assumed in the Woodson procedure, work sheets can be printed as shown by Form I, with the voltages 120 degrees apart. Actual cases, deviating from this balanced condition, will result theoretically in failure of the three perpendiculars to meet in a point. Experience with thousands of field cases on 3-wire delta systems shows the degree of such failure is not significant. Effects of voltage unbalance on types of services other than 3-wire delta are discussed later.

When the phasor diagram has been constructed, it is inspected for agreement with previously obtained data. For example, if, during the stator tests, one stator of the meter failed to rotate the disk the diagram should show either a nearly 90-degree relation between the respective stator potential and current, or a negligible a or c current. In case of doubt the stator tests are repeated. Ampere values in the data table are compared with diagram current phasor magnitudes. Diagram current phasors should theoretically add to zero and they are visually checked for this total. Wattmeter readings are totalled and compared to the recorded amount from the disk constant registration test. Observe that the diagonal addition, Fig. 2(A), table measurements of +30 *WMA*, +60 *WMB* and +60 *WMC* total 150 watts.

The tester next compares his judgment of load phasors with the diagram. In case of doubt about the load-diagram match, a repeat test with known load or known load change, or a visual checking of transformer connections are considered. If the operator is not satisfied by these extra measures, decision on the match of diagram phasors with load phasors is delayed until a repeat test under favorable conditions can be made.

REGISTER GEAR, BURDEN TESTS, AND OTHER FACTORS

To verify correct meter billing, the meter disk constant tested previously, the register gear ratio, and the transformer ratios are independently checked by the operator. The register ratio proving is delayed until this time in order not to disturb the "As Found" operation of the meter. The register is now removed and its performance determined on a separate device. If a mechanical demand meter is included in the installation, disk revolutions are matched against pointer deflection. The timing motor, if action is suspect, is tested for a period of one minute. The pusher pointer reading of a thermal demand is compared against the watts load.

A few installations have kilowatt and kilowatt-hour totalizing features. This requires checking of all contact devices, relays, auxiliary demand devices, and impulse values, and establishing proper transmittal to the end indication.

Relays and other nonenergy metering devices connected to instrument transformers may prevent opening of their circuits and thus complicate connection verifications. The meter-multiplying constant is not verified until transformer nomenclature, recorded at convenience during the test, is later checked with office records of transformer-acceptance tests. At this time, test-accuracy records of the transformers are reviewed.

Many users incorporate a current-transformer burden tester in the Woodson kit. The most economical time to use this on transformer secondary and meter-current circuits *A*, *B*, and *C* is while the wattmeter table data are being obtained for currents *a*, *b*, and *c*, respectively. The burden-tester gives indication of many major defects in the instrument transformers, and it should detect most grounds in the meter and transformer connections. Any ground fault with approved standard circuitry not sensed by burden test, should cause only a small meter error.

Instead of a "kit," three separate meters, one voltmeter, one ammeter, and one wattmeter, are sometimes used to obtain the Woodson data. This does not change the procedure in principle but increases time of testing. Time of testing is important because loads change and the more quickly data are obtained the greater chance there is of all readings representing the same load, a requirement for a firm diagram.

In the author's company, the watt-hour meter constant and connection accuracy tests (Fig. 1) of an installation call for removal and replacement of the register. This in turn automatically demands, at a later date, a meter calibration test.

PROCEDURE SUMMARY

Assuming standard circuitry, normal balance of voltages, and a known load, but with the limitations as shown by examples described later, the Woodson procedure enables the operator to give quick and accurate decision on the correctness of transformer connections to the meter test switch. The major key to trouble in this statement is the term "known load." Procedures other than the Woodson are necessary to complete the verification of the installation by proving the meter and connections thereto from the meter test switch. Only one of these tests as described later is infallible.

Discussion of Limitations

In 1928, the relative simplicity which was required to prove instrument-transformer connections by the Woodson method facilitated judgment and made the procedure almost foolproof. However, two assumptions were necessary. Woodson stated that the load power factor should be known in order to be within 60 degrees of the diagram phasors, and that it should be known as to whether they are leading or lagging.[1] Knowlton also described limitations.[2] Kouwenhoven, in 1916, argued and Woodson in 1928 implied, that any industrial meter installation could be tested when only a 3-phase motor was operating, thus obtaining a balanced load with a lagging power factor capable of being estimated within the 60 degree requirement.[1,3] Accordingly, on 3-wire delta systems, with the exception of the 3-wire version of example 4 described later, the Woodson check in 1928, when the precautions its author cited were followed, could give extremely good security.

Today most energy takings cannot be reduced to a simple balanced 3-phase load because of their nature or because customers do not desire to stop production during normal working hours. Hence, in many cases the Woodson tester is forced to accept an "unproved" load for his test condition and thus falls short of making the firmest possible check. This is the most important factor in restricting the accuracy of the Woodson method.

Since 1928, 4-wire Y, 4-wire delta, and network systems have become increasingly prevalent. The Woodson kit was originated to be used in 3-stator check of a 3-wire delta system. The phasor diagram data obtained at the meter switch established the three currents at the meter switch. Nine wattmeter readings were required to determine these currents independently and afford a check to the operator's work. Correspondingly, to determine the correctness and verify the four current phasors of a 4-wire systems at the meter switch, 12 wattmeter readings are necessary, 16 desired; any fewer than these degrades the test.

The question probably arises as to why one more stator is needed for the kit than is required in the corresponding meter where Blondel's *N*-1 theorem is applied. The answer lies in the fact that the kit does not measure power or energy as an end result but measures the quantities needed to construct a firm phasor diagram. A firm diagram requires two intersecting perpendiculars to determine a current phasor, and a third to prove that the operator made no significant mistake in obtaining data and plotting it. This means that the kit must measure one current and three wattmeter quantities for each current phasor to be established. Establishing all currents on the phasor diagram from measured data permits examination of their vector addition to zero. This control is necessary to decrease the possibility of human error in constructing the diagram. Thus, if the neutral current of a 4-wire 3-phase Y is not constructed from wattmeter measurements but is obtained from vector graphical addition of the 3-phase currents of the diagram, it is impossible to ascertain complete soundness of the operator's work in making the diagram.

FOUR-WIRE Y

Because of the difficulty of utilizing the neutral-to-ground voltage for wattmeter measurements only $3^1/_2$ stators appear practical for use on a 4-wire Y system. This means twelve wattmeter readings obtained with four conductor currents and three voltages to the kit neutral. The lack of the four wattmeter readings obtained with the neutral-to-ground voltage means that the neutral-to-ground voltage phasor cannot be constructed and checked by inspection for consistency of measurements, reasonable phase position, and addition to zero with the 3-Y voltages.

FOUR-WIRE DELTA

It might appear that in using the 4-wire delta system there would be available four test currents and four or more usable potentials, making possible the construction of a diagram of phasors. However, actual metering practices applied to 4-wire delta may prohibit a good Woodson check. This depends upon the voltages and currents available at the meter test switch from the instrument transformers. If three primary conductor currents of a 4-wire delta are provided at the meter test switch, the fourth current can be measured by summation in the kit. If, in addition, three voltages from the potential transformers are available at the meter switch, it is possible to construct a diagram with three intersecting perpendiculars to determine each of the four current phasors. If the three available potentials are the conductor-to-conductor voltages, they form a closed polygon and their sum is zero; this affords vertical and horizontal checks to zero of measurements in the wattmeter table. If the sum of any three meter switch potentials is zero or nearly zero, normally this also assures wide phase-angle separation, a requirement for accurately locating the intersection of diagram perpendiculars. If the three potential points are the corner vertices of the delta, the kit potential coil voltages will be the balanced E_{an}, E_{bn}, E_{cn} Y of the 3-wire delta; see Form I (Fig. 1) and example 1 in Fig. 2(A). Metering potentials, of course, need not be these Y voltages, but in any case the metering potentials could be drawn on the work sheet form similarly to the 2-stator metering voltages E_{ab} and E_{cb} of Form I. In actual practice, with a 3-stator meter or with one form of the $2^1/_2$ stator meter, if the three delta-vertice potential points and 3-conductor currents are available at the meter test switch, a satisfactory diagram can be established. However, another 4-wire delta metering method uses a 2-stator meter with totalized sensing of two conductor currents by the use of one stator. This generally affords only three potentials and one separate conductor current at the meter test switch, forbidding a diagram or a single, simple stator test. Consequently, ability to prove connections should be a consideration in choosing a meter.

SYSTEM VOLTAGE SUPPLY UNBALANCE

Unbalanced supply voltages were not discussed in the NELA 1928 Woodson paper.[1] There are two types of voltage unbalance, one caused by loading and one by design of the voltage supply. The latter type is exemplified by the 4-wire delta, by the 2-phase 3-wire supply, and by the 120-120-208-volt service where only three wires of the 4-wire network system are used. As with Woodson's original 3-wire delta arrangement, all other services require that voltage phasors be assumed for diagram plotting. As previously stated, these voltage phasors are the nominal voltages existing across the identical potential coils of the kit. Woodson plotted "kit" watts along the three balanced Y kit voltage phasors E_{an}, E_{bn}, and E_{cn}, which were assumed from connection of the kit to a delta system. This was permissible only because the potentials of his three wattmeters were equal.

Explanation of the use of watts as a phasor quantity was not given in the Woodson paper.[1] This explanation is required, however, to understand application of the Woodson method to unbalanced voltage services. Working backward from the completed phasor diagram example 1, Fig. 2(A), it will be seen that the distance *n* to *r* is actually the watt component $I \cos \theta$ of I_c on the voltage phasor $-E_{an}$, where θ is the included angle between $-E_{an}$ and I_c. This length

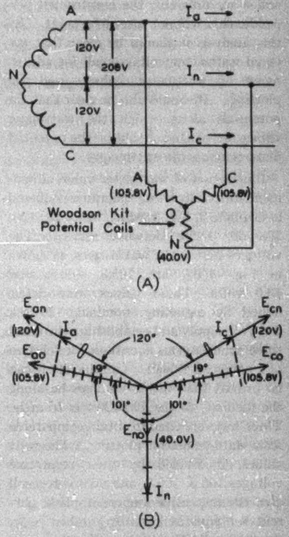


Fig. 3. Example 2: Service, 3-wire network, 120-120-208 volts, type meter, 3-wire 3-phase 2-stator, Woodson readings 240 watts

	A	C	N
Amperes	1.0	1.0	1.0
WMA	+100	− 80	−20
WMC	− 80	+100	−20
WMN	− 20	− 20	+40
Total	0	0	0

(A)—Three-wire network supply

(B)—Woodson phasor diagram

nr is then a measure of amperes. It will be remembered, however, that nr was also by construction $+30$ watts along $-E_{an}$. Hence, from this analysis, $nr=+30$ watts $=-E_{an}\,I\cos\theta=I\cos\theta$ amperes, or watts are equated to amperes. The explanation of this fallacy lies in the fact that watts are not a phasor quantity and for this specific purpose $+30$ is a numeric scale distance obtained by multiplying $I\cos\theta$ in the wattmeter by the numeric $-E_{an}$. Now, to be precise, I_c is established from perpendiculars erected at the end of the three scaled watt components of I_c on the three phasors, $-E_{an}$, $-E_{bn}$, and E_{cn}. The watt components of I_c are determined by the load power factor and the load in amperes. They are multiplied by their respective voltages (numerics) in the wattmeters. It follows that if the watt table values are to have the same plotting scale, the wattmeter voltage multiplying factors must be equal. If the scales are different, the constructed perpendiculars will not meet in a point. By this analysis it should be clear that unequal wattmeter voltages call for adjustment of wattmeter values used for plotting. It should also be clear that the potentials along which the wattmeter values are plotted should be the potential phasors across the wattmeters.

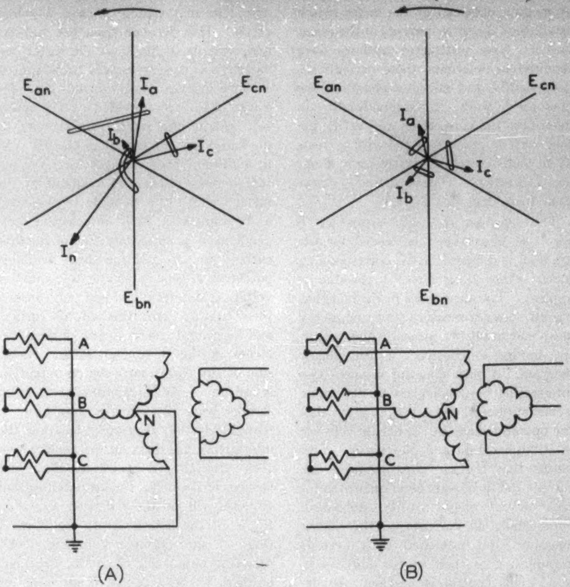


Fig. 4. Example 3: Service, 4-wire Y; type of meter, 4-wire 3-phase 3-stator

(A)—As found: Woodson readings 200 watts. Potential voltages in volts: AN, 126; AB, 213; BN, 122; BC, 212; CN, 125; CA, 217

	A	B	C	Total
Amperes	1.9	0.3	2.0	
WMA	+ 60	+20	−120	− 40
WMB	−155	−20	− 40	−215
WMC	+ 95		+160	+255
Total	0	0	0	0

(B)—As left: Woodson readings 195 watts. Potential voltages in volts: AN, 123; AB, 209; BN, 118; BC, 211; CN, 126; CA 215

	A	B	C	Total
Amperes	1.1	1.0	1.3	
WMA	+70	+35	−105	0
WMB	−80	+50	+ 30	0
WMC	+10	−85	+ 75	0
Total	0	0	0	0

Illustration of wattmeter value adjustment for constructing diagrams is offered in example 2, Fig. 3, which has a 120-120-208-volt 3-wire network service. The voltages across the wattmeters, as shown in Fig. 3(B), are 105.8, 105.8, and 40.0 volts. These values were determined by assuming nominal 120-120-208-volt supply and establishing the usual three simultaneous equations for solution of a balanced Y load. Graphically it can be proved that the Y voltages lie along the medians to the 120-120-208 triangle. Three ways are cited to obtain comparable scale watt values for plotting. The watt values, if divided by their respective voltages 105.8, 105.8, and 40.0 volts, will give the respective ampere in-phase currents or equalized plotting values. Or, the 40.0-volt, E_{n0} watt readings, namely -20, -20, and $+40$ watts, may be multiplied by 105.8/40.0. A simplification to avoid manual multiplication would be to incorporate the multiplication by means of different scale divisions on the voltage phasors printed on the work sheet form. This scheme is used in Fig. 3(B). Observe that the scale graduations for plotting watts from the data table along E_{n0} are 105.8/40 times as long as the scale along E_{a0} and E_{c0}. Actually this was readily accomplished by making a scale along E_{n0} 100/40 times a scale along E_{an} and E_{cn} and dropping perpendiculars from E_{an} scale markings to E_{a0} and from E_{cn} to E_{c0}.

Fig. 3(B) is a composite of the load phasor and the kit phasor diagrams. It is not correct in that when one diagram was superimposed on the other, points N and O, having a potential difference of 40 volts, were joined. This is permissible because, for the restricted purpose of the test no errors are made. However, the liberty taken permits easy construction and inspection of the meter test switch current and potential phasors in their proper phase sequence. Observe, as in the balanced delta and Y services, that current vectors add to zero on the diagram and are proportional to the measured values in the Woodson data table.

Similarly to the network system, if a

Vol.
Pt. 1
195

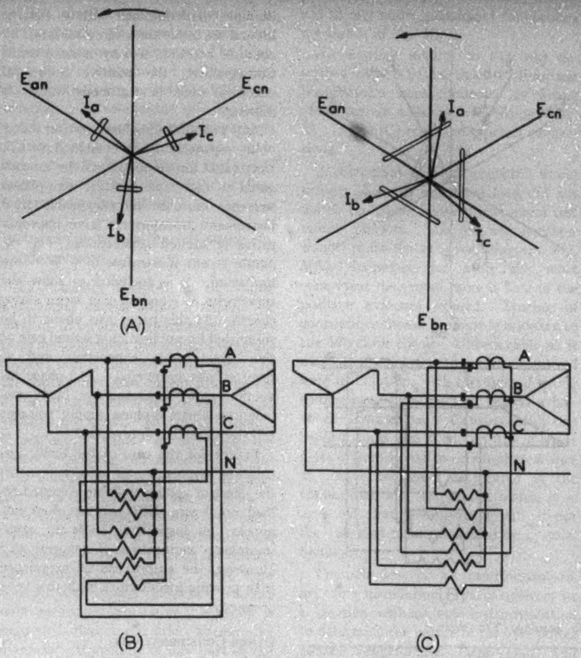


Fig. 5. Example 4: Service, 4-wire Y, type of meter, 4-wire 3-phase 3-stator. Potential voltage, in volts: AN, 120, AB, 208, BN, 120, BC, 208, CN, 120, CA, 208

(A), (B)—As found: Woodson readings 165 watts

	A	B	C	Total
Amperes	3.0	3.0	3.0	
WMA	+55	−20	−35	0
WMB	−35	+55	−20	0
WMC	−20	−35	+55	0
Total	0	0	0	0

(C)—As left: Woodson readings 60 watts

	A	B	C	Total
Amperes	3.0	3.0	3.0	
WMA	+20	+35	−55	0
WMB	−55	+20	+35	0
WMC	+35	−55	+20	0
Total	0	0	0	0

120-volt–120-volt–169.7-volt 2-phase 3-wire service is used, the Woodson nominal Y voltages are 89.4, 89.4, and 56.6 volts with 108.5 degrees separation between each pair of 89.4 volts, and the 56.6-volt phasors and 143 degrees between the two 89.4-volt phasors. This, of course, presumes identical potential coil impedances in the kit.

The 4-wire delta service has previously been analyzed in part. If metering follows Blondel's theorem, a review of proposed metering methods shows that the three Y potentials of the delta and all four primary currents may be sensed by the kit, permitting use of the three plotting potentials of Form I regardless of the particular voltages applied to the meter. Since Woodson's three required perpendiculars per current phasor depend upon the potentials and currents available at the meter test switch and not at the meter, even if Blondel's theorem is not used in selection of the meter it still may be possible to construct a firm diagram to represent test switch quantities. In the latter case, whether a special work sheet form is required would depend upon the particular conductor potentials and currents available to the operator.

Regardless of the service, experience and the use of appropriate Woodson work sheet forms make dissymmetry of potential supply a minor matter as far as the mechanics of obtaining a phasor diagram are concerned.

LOAD VOLTAGE UNBALANCE

Unbalanced voltages caused by load should always be considered a danger sign, although only occasionally are they liable to result in improper verification. A review of several hundred metering installations on 4-wire Y systems was made to help determine the practical influence of load voltage unbalance in proving connections for this type of service. System neutral conductor-to-ground voltages were found to be as high as 4 volts in congested areas and as high as 14 volts on long rural lines. The study indicated that the danger resulting from unbalance was slight where approved primary distribution connections were made. However, example 3 (Fig. 4) with only a few volts unbalance, represents a case of a nonstandard distribution connection and emphasizes the need of vigilance in testing. The study also resulted in the conclusion that for 3-wire and 4-wire Y services, printed work sheets such as Form I, with 120-degree spaced voltage phasers and normally obtained Woodson data could be used, if operators had a high degree of competence, for constructing phasor diagrams without danger of approving wrong connections.

At this time, the reader is requested, before reading further, to decide if the phasor diagram of Fig. 4(A) represents correct connections for single-phase and 3-phase induction motors. The as-found meter connections of Fig. 4(A) are correct, as the voltages and currents in the supply meter side of the transformer were as shown. In contrast, the customer's load phasors were those of Fig. 4(B). The difference is caused by the questionable neutral-to-ground connection on the Y primary of the supply transformer, together with the unbalanced supply voltage. Overloading of the distribution transformer with zero phase-sequence cur-

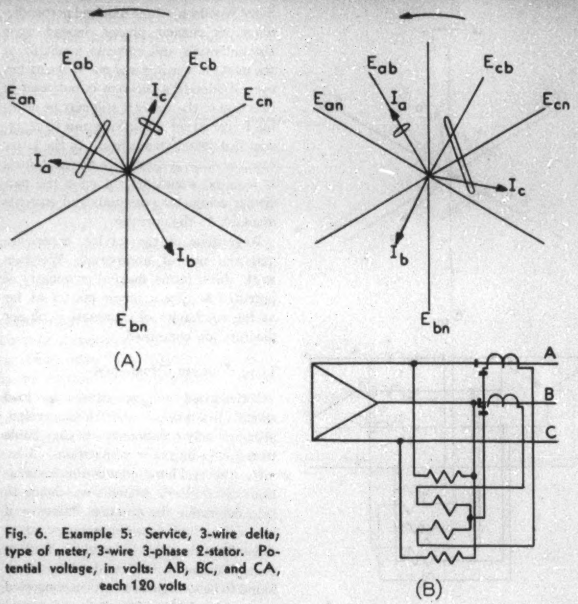


Fig. 6. Example 5: Service, 3-wire delta; type of meter, 3-wire 3-phase 2-stator. Potential voltage, in volts: AB, BC, and CA, each 120 volts

(A)—As found: Woodson readings 175 watts

	A	B	C	Total
Amperes	1.8	1.8	1.8	
WMA	+60	−65	+ 5	0
WMB	−10	+65	−55	0
WMC	−50		+50	0
Total	0	0	0	0

(B)—Alternate: Woodson readings 170 watts

	A	B	C	Total
Amperes	1.8	1.8	1.8	
WMA	+65	− 5	−60	0
WMB	−65	+55	+10	0
WMC		−50	+50	0
Total	0	0	0	0

rents results. Thus, although meter connections are correct, transformer and supply line currents are increased and the customer is charged with unnecessary transformer losses. When this questionable connection was removed the excessive transformer losses did not appear and only the customer's normal load was measured by the meter.

Observe the reverse power of meter stator *B* in the as-found table of Fig. 4(A). Note that the actual customer's current phasors or the as-left Fig. 4(B) phasors are the as-found phasors minus one third of the neutral or zero-phase-sequence current. There is temptation for the installer or a verification tester to make all stators "run" forward, thereby establishing an incorrect connection. Observe that if the voltages were balanced at the time of testing, the questionable connection would not be discovered. The vertical table totals under *A*, *B*, and *C* should not theoretically be zero but they were within the accuracy of measurements. Wattmeter table data in this example were obtained with a 3-stator kit and the total values were established by operator addition, not by measurement. Consequently, the positive horizontal check to zero of wattmeter values is missing.

Any potential unbalance of over a few volts, especially when combined with a watt table unbalance should be investigated to determine whether zero-phase-sequence currents are responsible or if instrument transformers have improper ratios or faulted connections. The example is not illustrative of a Woodson limitation. It is included to show the superiority of Woodson over meter stator checks. As this paper was about to be submitted for publication a second case of this questionable connection and I_b reversal was uncovered. A 3-phase induction motor operating on badly unbalanced 3-wire 3-phase supply voltages will show a similar I_b reversal.[4]

Except for the case of the badly unbalanced supply to an induction motor, the effect of voltage unbalance caused by load on 4-wire delta services does not appear, by inspection, liable to cause improper "verification of connections." However, the author has no experience with proving installations with this type of supply.

OTHER CONSIDERATIONS

Instrument transformers connected 3-wire Y-primary, 4-wire Y-secondary, may give some 12% overvoltage on the secondary as a result of the suppressed third-harmonic magnetizing current. This overvoltage is not indicative of a wrong connection but it does present need for matching knowledge and reasoning by the Woodson operator.

In increasing instances the addition of static capacitors to power systems has caused difficulty in determining whether the power factor is lagging or leading. Especially is this true of single-phase capacitors where primary capacitor phase and corresponding secondary metering phase cannot readily be paired. For illustration, single-phase welders with concealed internal capacitors may cause disturbing phasor diagrams especially on unbalanced system voltage supplies. For the effect on a balanced 3-wire delta service observe the example on Form I (Fig. 1). Likewise, the practice of ganging "leading-power-factor" fluorescent lights on one phase and "lagging" on another presents a hazard to proper verification. Up to 30% of third harmonic and its multiples in fluorescent-lighting phase conductors can also cause confusion to the operator. Saturable reactors used for

Vol.
Pt. 1
1959

voltage control can cause current to be practically all harmonic with confusing results to an operator. The interconnection of systems and customer generation with tie-line metering installations, together with standards of operation which do not permit Woodson operators to open such interconnections, again create the hazard of unknown load power factor. For illustration of some of these factors examples follow.

EXAMPLES OF WRONG TRANSFORMER CONNECTIONS

All examples described in this paper were found on the author's system. A total of 13 incorrect transformer connections in 5,000-meter installations were uncovered during the years 1947–1957. In the wattmeter tables, except for example 3, Fig. 4(A), nominal rather than actual voltages are given. Excepting example 6, wattmeter values in the tables are rounded out to give horizontal and vertical checks to zero, whereas in practice there may be minor discrepancies.

Example 1, Fig. 2(A), shown previously, represents the case of a department store installation with cold cathode fluorescent and incandescent lighting. There were several single-phase induction motors with associated capacitors. Now, it is suggested that the reader answer the question: If opening of any part of the load in customer selling hours is objectionable, how many operators would accept the Fig. 2(A) phasor diagram as proof of proper connections? The operator in the actual case had reason to suspect fluorescent lighting and capacitors. He returned after normal work hours and made a test with only incandescent lamps as a load. This led to discovery that *A* current transformer had been installed with reverse polarity in *B* conductor; see Fig. 2(B). Fig. 2(C) represents correct connections.

Observe that the phasor diagrams for both the wrong and right connections (Figs. 2(A) and (C) respectively) give correct energy registration, or 150 watts. However, a light load is depicted and an increase in loading will cause the as-found listing in Fig. 2(A) connections to cause under-registration. Tests under light load conditions where fluorescent lighting or capacitors are a major part of the loading should be suspect.

Examples 2 and 3 have been discussed. Example 4 represents load-comprised induction motors, capacitors, and incandescent lighting. Fig. 5(A) is the phasor diagram at the meter test switch. The reader may decide as to the correct connections. This case demonstrates a hazard to any audit. As-found connections are shown in Fig. 5(B).

An example similar to this was described in the NELA article.[1] All current transformers were connected to wrong meter current coils and with wrong polarity. Fig. 5(C) represents correct connections.

A competent operator probably would suspect that the estimated load did not match the power factor of the meter test switch phasors. The average operator would probably approve diagram Fig. 5(A). However, for even the most competent meterman there is lack of the positive evidence desired. Tracing of connections should be made in addition to the Woodson check. Unbelievable as it may seem, 2 of the 13 errors uncovered were installations with this type of wrong connection. Observe that this is a balanced 3-phase load. Consequently, this type of fault can occur on any 3-phase delta system as well as on a 3-phase 3-wire Y. Since the error is located in the connections and not in the meter, it is expected that the operator will accept the 165-watt measurement as being satisfactory.

Example 5 covers an intercompany-tie metering installation in rural territory at a location without any instrumentation to determine the nature of the load power factor. Single-phase farm load and single-phase capacitors were evident. Energy could flow from both ends of the line. Would Fig. 6(A) be accepted as representative of the load and proof of proper connections? The reader may wish to decide. After Fig. 6(A) was determined, it was requested that one generating station have its excitation increase. The component change in current at the test location was thus known to be leading and a repeat test gave more evidence that the as-found diagram was correct. In this case, because of the demand for more integrity of decision, all visible connections were checked. As the diagram appears reasonable, Fig. 6(B) representing a wrong connection, is included to show how easily it could be approved. Although the error of the Fig. 6(B) case is only the difference between 175 and 170 secondary watts, any increase in load will probably cause more lagging current and a percentage decrease in amount of proper registration.

Example 6 (Fig. 7) shows the superiority of a "Woodson" over a meter stator check. The load was incandescent lighting, induction motors, and no capacitance. A meter stator tester may attribute "backward" rotation of stator *A* to low power factor. Observe, also that a meter installer may reverse I_a leading to an improperly verified installation. The phasor diagram showed an obvious error. A ground was found on the secondary wiring of the current transformers. Inspec-

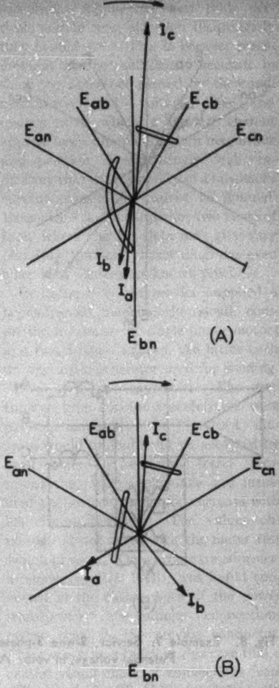


Fig. 7. Example 6: Service, 3-wire delta; type of meter, 3-wire 3-phase 2-stator. Potential voltage, in volts: AB, BC, and CA, each 120 volts

(A)—As found: Woodson readings 115 watts

	A	B	C	Total
Amperes	1.0	0.8	2.3	
WMA	−30	−20	+ 65	+15
WMB	+65	+55	−155	−35
WMC	−35	−35	+ 90	+20
Total	0	0	0	0

(B)—As left: Woodson readings 120 watts

	A	B	C	Total
Amperes	1.4	2.0	2.4	
WMA	+25	−60	+40	+ 5
WMB	+25	+50	−85	−10
WMC	−50	+10	+45	+ 5
Total	0	0	0	0

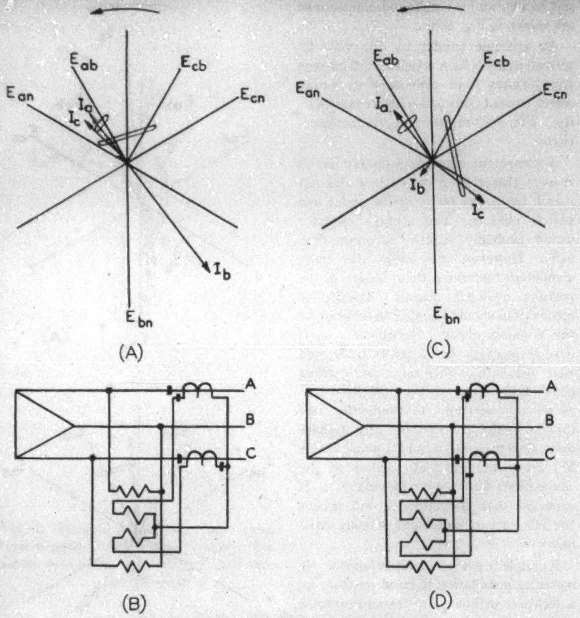


Fig. 8. Example 7: Service, 3-wire 3-phase delta; type of meter, 3-wire 3-phase 2-stator. Potential voltage, in volts: AB, BC, and CA, each 120 volts

(A), (B)—As found: Woodson readings 270 watts

	A	B	C	Total
Amperes	1.7	3.5	1.3	
WMA	+100	−220	+100	−20
WMB	− 90	+190	− 80	+20
WMC	− 10	+ 30	− 20	0
Total	0	0	0	0

(C), (D)—As left: Woodson readings 155 watts

	A	B	C	Total
Amperes	1.5		1.5	
WMA	+100	+10	−110	0
WMB	− 90	+15	+ 70	−5
WMC	− 10	−25	+ 40	+5
Total	0	0	0	0

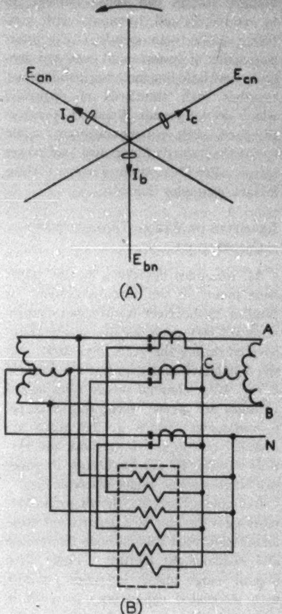


Fig. 9. Example 8: Service, 4-wire Y, type of meter, 4-wire 3-phase 3-stator. Potential voltage, in volts: AN, BN, and CN, each 120 volts, AB, BC, and CA, each 208 volts

As found: Woodson readings 300 watts

	A	B	C	N	Total
Amperes	10.0	5.0	10.0	5.0	
WMA	+120	−30	− 60	−30	0
WMB	− 60	+60	− 60	+60	0
WMC	− 60	−30	+120	−30	0
Total	0	0	0	0	0

tion shows that I_b was advanced in phase position by the fault, and I_a retarded. If the ground fault had been at another location these effects might have been reversed with the result that the meter disk would show "forward" rotation for each individual stator. Hence, in this assumed version, if tests for grounds were not made there would be no significant clue to signal either meter stator or Woodson tester.

Example 7 shows an unbalanced induction motor and incandescent lamp load, again included to show the superiority of a Woodson over a meter stator test. Assume that reliance for verification was placed upon a stator check to determine correctness of transformer connections. Then, in the as-found phasor diagram, both stators will run forward, indicating a reasonable proof of proper wiring. On the other hand, trouble is clearly indicated by the three current phasors determined with the kit. As-found I_c of Fig. 8(A) and (B) was reversed and is corrected in as-left listings in (C) and (D). This unbalanced load with low power factor and the loads of examples 4 and 9 are exceptions in that fullload registration is greater with incorrect connections.

Example 8, in Fig. 9(A), is the result of a 3-stator Woodson test of a 4-wire Y load of incandescent lighting. The reader may decide as to the proper connections.

Vol.
Pt. 1
1959

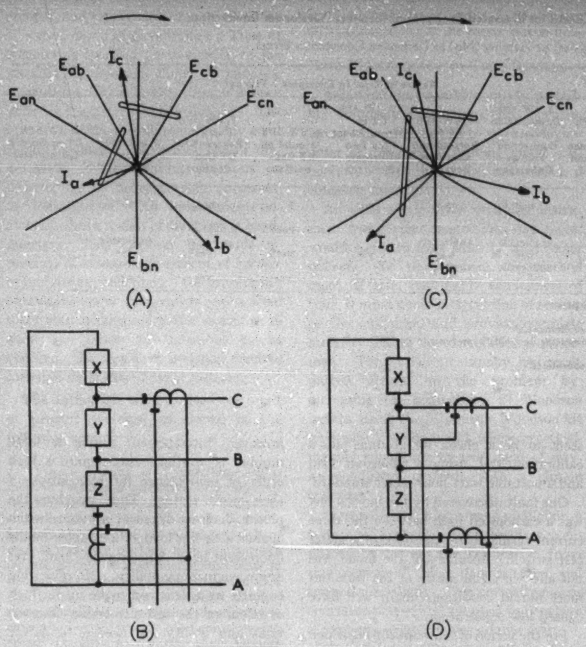


Fig. 10. Example 9: Service, 3-wire 3-phase delta; type of meter, 3-wire 3-phase 2-stator. Potential voltage, in volts: AB, BC, and CA, each 120 volts

(A), (B)—As found: Woodson readings 255 watts

	A	B	C	Total
Amperes	1.2	2.6	2.4	
WMA	+65	−175	+110	0
WMB	+30	+135	−165	0
WMC	−95	+ 40	+ 55	0
Total	0	0	0	0

(C), (D)—As left: Woodson readings 230 watts

	A	B	C	Total
Amperes	2.3	2.2	2.2	
WMA	+ 40	−155	+115	0
WMB	+120	+ 40	−160	0
WMC	−160	+115	+ 45	0
Total	0	0	0	0

This is a definite hazard case. I_b is actually I_n as the current transformer is in the neutral. I_n and I_b are identical. Observe that both I_b and I_n are required to balance the vector sum of I_a and I_c. How many metermen using a Woodson kit or making a meter stator investigation would uncover this error? An alert Woodson operator realizes the danger of a possible misplaced current transformer, especially when inspection shows that the neutral current approximates a phase current, and as a consequence he performs the indicated additional checking. Although there is no registration error at the condition shown, normally with a more balanced load the meter would measure only partial energy consumption. Again, the watt table verticals add to zero, although theoretically they should not. This is because measurement was not sufficiently accurate to show the difference caused by the small neutral-to-ground voltage. The frequency with which a neutral current approaches a phase current in magnitude and position is surprisingly high, indicating that the possibility of a misplaced current transformer cannot be ignored. Examples 4 and 8 constitute two types of fault where the kit data may give very plausible "proper" phasor diagrams even when load phasors are known precisely.

In example 9, the service supplied a typical small machine shop load, comprising a number of single-phase motors and two 3-phase motors, the latter each driving a d-c generator used for welding. There were no capacitors. The two through-type current transformers were pole-top mounted. The Fig. 10(A) diagram would undoubtedly be accepted by most testers. The actual tester questioned the phasor unbalance and measured the service conductor currents with his clip-on ammeter. The values obtained did not agree with the meter test switch phasors. One current transformer, as shown by Fig. 10(B), was found connected in the closed delta of the power transformer low-voltage connections rather than in the service conductor. There was a possibility in this case that a careful visual check of connections from the ground might have disclosed the fault. Hidden in some cubicle, such a wrong connection would probably defy detection. This load had a very low power factor and caused the true energy to be less than the as-found metered quantity. Figs. 10(C) and (D) show correct connections.

In Table I eight of the examples just described involving wrong transformer connections were subjected to several of the more popular tests. Opening-line or primary wires was not included as it is no longer considered practical. The Kouwenhoven check requires a nearly balanced load or the test becomes complicated for 3-wire cases. It is not applicable to four wires. The Ackerman and opening voltage leads checks failed almost completely.

Meter Checks

Internal meter connection faults have not been adequately discussed. Examination of modern polyphase watt-hour and watt-hour demand meters shows that internal terminal reversal of potential coils is possible. Grounds are possible. Current coil reversal is not probable. Coil

Table I. Popular Checks for Illustrated Examples of Incorrect Transformer Connections

Ability (Yes) or Inability (No) to Determine Connection Errors

Nature of Check, Conditions	Nature of Error in Examples 1 Through 9							
	1, 3W-3Ph CT in Wrong Phase Plus Current Reversed	3, 4W-3Ph Questionable Primary Connection: Voltage Unbalance	4, 4W-3Ph 3 CT's Rotated and Currents Reversed	5, 3W-3Ph CT in Wrong Phase Plus Two Currents Reversed	6, 3W-3Ph Ground on CT Connections	7, 3W-3Ph Current Reversed in CT Connections	8, 4W-3Ph CT in Neutral	9, 3W-3Ph CT in Supply Transformer Delta
Opening potential coil leads: necessary to know if power factor is more than 50%	no	no*, †	no*	no	‡	no	no*	no
Ackerman: unbalanced load not more than 25%; necessary to trace leads	no	not applicable	not applicable	no	yes§	§	not applicable	no
Kouwenhoven: leads must be traced; load should be balanced	yes¶	not applicable	not applicable	yes	yes¶	yes¶	not applicable	?
Woodson: knowledge of load connections	‖	‖	‖	yes	yes	yes	yes	‖

Note: CT = current transformer; 3W = 3-wire; 3Ph = 3-phase.

* Two coils are opened to give single phase operation.

† If installer had reversed phase *B* on forward rotation of element *B*.

‡ Shows wrong meter rotation: may cause meterman to reverse I_a rather than remove ground.

§ Questionable rotation of meter; questionable answer.

¶ Test not applicable, however, if unbalance of 25% is considered.

‖ Yes, but only under conditions described in text.

interchange does not seem possible. For graphic meters both coil reversal and interchange are possible. As with the instrument transformers, the possibility of faults in meters combined with tests made with relatively small currents in one phase, capacitor power factor or voltage control and 4-wire services have made meter stator tests, using the customer's load, more vulnerable to error.

In the 5,000 cases covered by this report only one internal meter-wiring fault was discovered: a reversed potential coil in a graphic meter. However, if the ground of example 6 (Fig. 7) had occurred in the meter after installation, instead of on the transformer secondary, probably no usual meter test with customer load would show the fault. A transformer burden test, used in many companies, probably would uncover the defect. The diagram, from data taken at the meter switch, would normally show the incorrect as-found phasors. However, this assumes that the transformer ground is on the commoned side of the current coils. Nonstandard meter test switch wiring (which does occur) might cause an as-left version of the diagram and result in improper verification. If a reversed meter coil is substituted for the reversed transformer leads of example 7, the diagram will show the as-left arrangement rather than the trouble indicated by the as-found relation. Only the registration discrepancy of 155 Woodson watts against 270 customer-meter watts would give a clue to the error. Another load condition might decrease this difference sufficiently to afford no clue.

If it is conceded that very infrequently the internal wiring of a meter may be misconnected, the best "field" defense would seem to be a check for grounds plus a calibration test using a phantom load and a portable watt-hour meter standard.

One fault discovered by the burden test was a carbonized path between the three current terminals on the back of a meter test switch. Accuracy of the meter was not affected significantly at the time but more humid conditions might well have caused low registration.

For the period of ten years during which tests of 5,000 installations uncovered the 13 wrong transformer connections described, 13 billing constants were also found to be in error by reason of the verification checks disclosing wrong multiplying constants.

SUMMARY OF FIELD FINDINGS

The examples described herein incorporate the limitations of Woodson or other procedures and are included to demonstrate how wrongly connected meters can be "verified." Eight of the 13 incorrectly connected meters were found obviously wrong by casual inspection of the phasor diagram. Diagrams of examples 6 and 7 fall in this category although, as explained before, meter stator tests would probably not uncover the errors. The cases that demand serious consideration are those similar to examples 1, 4, 5, 8, and 9 where routine of "correct" installations is likely to dull the alertness of the operator.

Some advocates of the Woodson procedure consider that only limited competence is needed in field testing. These advocates consider that the responsibility for decision of phasor-diagram accuracy should rest with a test-reviewing engineer. They have a strong argument for all but a few per cent of installations. These few, however, interspersed among a large number of routine cases require a high order of competence for recognition of their special status. Thus, matching the phasor diagram taken at a meter switch against a tie-line load, or a complex-motor fluorescent-light capacitance load, requires engineering ability. No reviewing engineer at a desk can make up for lack of talent at the test site where observations and ability are needed to detect clues denoting a wrong connection. Sometimes, and perhaps for good reason, installations are not "standard" or present particular difficulties. Relays and other nonmetering devices connected to instrument transformers may forbid the opening of circuits. Meter-type application may be questionable. Local requirements and/or other demands of policy may make problems for the tester. Wiring or mountings may be inferior. This requires operators preferably with wide experience, sound judgment, tactful personalities, and dedication to maintenance of accurate metering.

Field operations using the Woodson method are considered economically justified in comparison with costs of other methods of proving connections. In the author's company the operator is furnished with an automobile, a kit weighing 36 pounds, associated tool and lead box weighing 23 pounds, an 8-pound register gear ratio checking device, a clip-on ammeter, and a telescope for reading pole-mounted transformer nomenclature. The 5,000 installations were checked in the past ten years at a rate of over four per day. The operation included travel from a centralized location, the auditing of complex totalizing installations, check of meter constants and transformer accuracy

Vol.
Pt. 1
1959

records, and report writing, as well as the Woodson check of connections. Time at a metering location for over 95% of the installations average about 45 minutes. Recently a kva metering test has been added to the tester's responsibilities. The operator, at the installation, quickly completes the meter test switch phasor diagrams as well as making decisions on connections as being correct, incorrect, or inconclusive. An independent engineer makes a review of diagrams at headquarters. This engineer has made no reversal of a tester's judgment of correct or incorrect connections. Geographically operations cover four states and a local meterman accompanies the operator to show the latter the locations to be checked. In turn the meterman receives instructions in this phase of metering.

The author has tried to show the merit of present systems, as shown in the following summary.

1. *Visual Tracing of Connections:* This method requires a minimum of equipment and technical competence. It is subject to such handicaps as pole-mounted transformers requiring the use of ladders, or linemen as checkers working with a meterman on the ground, transformer connections in conduit where splices of differently identified conductors are possible, fading and discoloration of conductor insulation identification, color blindness of tracers, mechanical tracing errors, inability to discover short circuits and grounds, physical hazards, and wrong nomenclatures. The human element is judged unequal to the task of approaching the required order of effectiveness where this procedure is used. The visual tracing of connections is generally supplemented with item 2.

2. *Meter Stator Tests:* A check of Table I shows that meter stator tests using the customer's load are deficient in the selected examples. Knowlton presents a comprehensive table of 3-wire 3-phase service limitations.[4] Any method which does not determine approximate magnitude and phase relationship of phasors is certain to fail too often. The meter stator test procedure is sometimes used as the sole means of verifying transformer and meter switch connections.

3. *Woodson Method:* Originally, in 1928, with only 3-wire services and ability to limit loads to balanced 3-phase known impedances, the procedure was almost foolproof. Since this procedure incorporates item 2 and, in doubtful cases, may occasionally use item 1, it readily merits top rank for integrity. However, today, the 3-stator kit for proving 4-wire services is not adequate, and the required determination of load phasors occasionally exceeds the capabilities of testers.

4. *Meter Calibration and Ground Faults:* This is a procedure which supplements any of the foregoing three examinations with burden and/or ground tests and a calibration test using a phantom load and portable watt-hour meter standard which will eliminate any significant errors in the meter or its connections to the meter test switch.

NEW PROCEDURES AND TEST DEVICES

Assuming the soundness of the preceding summary of findings, the problem facing the industry is to gain a higher order of integrity in proving meter connections. This paper does not provide a complete answer.

An ideal method for verifying instrument transformer connections to a meter would compare load phasors with meter phasors. To be practical the method must be safe, reasonably economical of time, it must avoid interruption of service to the customer, and provide complete integrity within the capabilities of metermen. The author's company has made partial attack on the problem by: increasing the competence of Woodson testers, using the 3½-stator Woodson kit on 4-wire Y services and, where practical, making magnitude measurement of service conductor currents with a clip-on ammeter.

Increasing the competence of Woodson testers is probably most easily accomplished by having the verification function performed by a few selected and trained specialists. This normally means centralization of the entailed work where a company operates over a large area.

The 3½-stator Woodson kit has been used by the author's company for 4 years. In many cases the fourth current it determined gave desired security. By means of this fourth current the table data can be checked horizontally and the phasor currents on the diagram can be added to zero. This proves that most of the data are consistent, a required step in the procedure.

Other means of adding to the integrity of checking have been tried. The recently introduced 15-75-150-750-ampere hook-on, transformer-coupled ammeter insulated for 8.7 kv to ground is a useful tool to measure load-phase currents when such a check is demanded and where conductors are accessible. This permits a magnitude comparison of meter switch phasors with service conductor currents. This instrument was also used in an experimental test to match load phasors against meter switch phasors by means of a cathode-ray oscillograph and an electronic switch. The corresponding meter switch and load phase wires were readily paired by observing simultaneously different combinations of a load current and a meter switch current. This resulted because the screen normally showed two waves either in phase or nearly 120 degrees out of phase. Phase sequence of the load was established by noting the left-to-right order of two load phase currents on the screen. Subsequently, the meter switch potentials were paired with the potential-transformer primary currents obtained with the hook-on ammeter. This required knowledge of the power factor of the input to the potential transformer and had the disadvantage of working with high harmonic content from the hook-on ammeter. Nevertheless, reasonable proof of connections was considered to be obtained since, by visual inspection, current and potential transformer primary connections were correct, and it followed that correspondingly paired meter switch currents and potentials must also be correct.

This method avoided the opening of any connection except at the meter switch. This is desirable as disconnecting wiring during tests presents the hazard of wrong reconnection. However, if the secondary terminals of the potential transformers are opened, an oscilloscope indication is obtained free of harmonic content and phase displacement, making pairing of waveforms a much simpler task. The proper working combination or pairing of voltages and currents could also be checked by opening a specific potential transformer at the secondary terminals and determining if the corresponding potential has disappeared at the meter switch. With sufficient output from the clip-on instrument, phasor-diagram data of the primary system could be obtained and the resulting diagram oriented and compared with the diagram at the meter switch.

The practicability of direct sensing of load phasors as described depends upon physical accessibility to the primary system. For this reason a large proportion of installations will not yield to this solution. Better sensing devices for current in, and voltages on, the primary conductors would also make the test more practical. The electronic wattmeter in combination with a workable Hall effect sensing device might possibily furnish an answer. Again, because the original Woodson test is faster and more convenient, the extended method just described is considered to have practical merit only where the normal procedure has failed to give conclusive proof of connections.

Rudd in 1936 proposed using Lissajous figures obtained from a cathode-ray oscillograph to construct a diagram of phasors at the meter switch.[5] Actually the publication title is limited to current transformer connections, and only an ammeter was used in addition to the cathode-ray oscillograph. There appears to be no reason why potential transformer connections could not be included. Essentially, the proposal substitutes the oscillograph for components in the Woodson kit, but the results obtained cannot have the exactitutde of the volt-ampere-wattmeter procedure nor does it provide easily recorded data for engineering review. Neither does it incorporate all the self-checking features used to prove integrity of the Woodson data and diagrams. The same knowledge of load phasors is required as with the Woodson method.

The Woodson operator could readily make acceptable separate stator registration tests of the meter with his kit equipment and the addition of a suitable resistor to obtain a known current coil loading. This would avoid errors caused by using an inadequately known customer load. The resistor would be inserted in potential supplies to obtain the test current. For example, potential AB could thereby furnish current for the corresponding current coil A where voltage AB and current A constitute torque requirements for a meter stator. This assumes that meter test switches are installed and a Woodson check is desired before the first customer billing. As yet in the author's company this test has not been made part of connection verification procedure.

Conclusions

1. Experience indicates that over 99% of connections in instrument-transformered metering are originally installed correctly. Of the errors made, the percentage detected by a connection verification test will depend upon the type of wrong connection, the load, the test procedure, and the abilities of testers. Thirteen wrong connections in a check of 5,000 installations were found on the author's system. The embarrassing question presented is, "How many were not found?"

2. A basic difficulty lies in the fact that the test load for proving transformer connections must be the customer load, a quantity which cannot always have its phasors adequately known. This feature of unknown load phasors has increased in the past 25 years because dissymmetrical voltage services, leading power factor and unbalanced loads, and occasions where a tester has been unable to command a simple test load, have all become more numerous. Consequently, the possibility of verifying an incorrect connection has increased. A survey of technical publications shows an absence of proposals to modernize 1928 practices to counter this added vulnerability.

3. Stator and other popular meter tests made at the meter test switch, using the customer's load, are insufficient to give required security either against faults in the meter or in connections to instrument transformers.

4. The Woodson method for finding faults in transformer connections to the meter test switch is considered superior to other procedures. With highly competent operators most errors in connections will be discovered.

5. Separate and combined stator calibration tests made at the meter test switch, using a phantom load and portable rotating standard with addition of a test for grounds will give practically foolproof indication of faults in the meter.

6. A better method of connection checking is a need, and should be a challenge, to the industry. As an example, one procedure would compare oscillographs of load phasors with meter test switch phasors. To be practical, a connection verification method must be safe to personnel and reasonably economical of time; it must avoid interruption of service to customers, and provide complete integrity of action.

References

1. National Electric Light Association Meter Committee Reports. *Publication no. 238-31*, Federal Printing Company, New York, N. Y., 1928, pp. 28–44.

2. Electric Power Metering (book), **A. E. Knowlton.** McGraw-Hill Book Company, Inc., New York, N. Y., 1934, pp. 184–206.

3. A Method of Determining the Correctness of Polyphase Wattmeter Connections, **W. B. Kouwenhoven.** *AIEE Transactions*, vol. 35, pt. I, Feb. 1916, pp. 183–211.

4. Effect of Unbalanced Voltages on the Receiving Circuit and Watthour Meter, **A. E. Knight.** *NELA Meter Committee Reports*, National Electric Light Association, New York, N. Y., App. 20, Apr. 10, 1931, pp. 127–28.

5. A Method for Checking Current Transformer Connections to Multi-element Watthour Meters by Use of a Portable Cathode-Ray Oscillograph, **W. C. Rudd.** *Bulletin*, Edison Electric Institute, New York, N. Y., Oct. 1936, pp. 425–27.

◆

Discussion

Thomas D. Barnes (Westinghouse Electric Corporation, Raleigh, N. C.): The author has done the industry a valuable service in his comprehensive paper. In complicated meter testing such as for connection verification, it is as important to know the limitations of a test method as it is to know the procedure.

The many variables entering into present meter installations such as dissymmetrical voltage services, leading load power factors, and unbalanced loads may well set up situations wherein the only characteristic that can be considered definitive is Blondel's theorem. The author shows, in Table I, the unresolved checks of the various test procedures currently used, and illustrates that no one method can give an unequivocal answer.

An important point is the advisability of employing highly competent technicians for checking out meter installations for integrity of connections. To the test results must be added the ability to trace connections quickly. Figs. 4, 7, and 8 are examples where connection and polarity tracing was either necessary or advisable to confirm the correctness of connections.

The recently introduced single-stator network meter is an example where another variable must be checked; that of phase sequence. Heretofore, phase sequence has been of importance only in var-hour metering.

Unfortunately for the meter tester, phase sequence of polyphase circuits are not standardized in the states as they are in Australia, for example, where all line conductors are coded by red, yellow, and blue markings.

The author mentions the need for some method of checking for proper connections of all installations. There may be a practical way but the need of establishing both the load phasors and meter test phasors without interruption of service makes the problem extremely complex. In the meanwhile, the author clarifies the proper use of available methods.

D. M. Smith (General Electric Company, Somerworth, N. H.): The paper is a timely and comprehensive yet understandable treatise on a complex and difficult subject. It shows that today's changing load pattern has markedly increased the hazard that substantial revenue may be lost by approval of incorrect meter connections. It is indeed fortunate that the total probability of such errors on a percentage basis is still very small.

In view of the difficulty of proving with complete assurance that the connections of an installation involving transformers are correct, it is highly desirable that the chances of incorrect wiring during installation be reduced to the absolute minimum, by adhering to standard connections wherever possible, among other factors. In general, the utilities are consequently highly reluctant, and rightly so, to adopt a metering scheme which involves nonstandard connections of instrument transformers.

To extend this principle a step further, it might well be used as an argument in favor of adopting the class *200* self-contained polyphase meter for 200-ampere services, in preference to the use of current transformers at this level, thereby reducing the hazard of incorrect connections. Such is already the practice of many utilities, although not necessarily for this reason.

This author makes an excellent case for the use of the Woodson method as the one best test, whether or not auxiliary tests are made, and it is to the credit of the orginator that, 30 years later, a better method has not been devised. It would be of interest to

Vol.
Pt.
195

know, in this connection, what percentage of utilities use this method, and what their experience has been in uncovering cases where the Woodson diagram was capable of misinterpretation by the average user.

The author mentions the use of a portable watt-hour meter standard for proving accuracy and integrity of meter and connections to the switch, together with use of a clip-on ammeter where practical, to check primary currents. At least one utility has taken a further step by using a polyphase portable meter standard in conjunction with clip-on current transformers, to get closer to customer's load. This of course is subject to limitations in accessibility of primary leads.

Since this paper highlights so successfully the shortcomings of the best-known methods of proving metering installation integrity, it is unfortunate that a better solution to the problem cannot simultaneously be proposed. It is to be hoped that this paper will focus attention on the need for further improvement in methods.

F. W. Warburton: Mr. Barnes and Mr. Smith are thanked for their comments. Their statements describe the problem with enhanced clarity. However, a possible misunderstanding concerning the need of phase sequence may result from Mr. Barnes' discussion. It is answered by the facts that the establishment of phase sequence is a component of the Woodson procedure and phase sequence indiction is part of the Woodson phasor diagram.

There was an absence of discussion from utility metermen. As reported in the paper there has also been an absence, during the past 25 years, of publications describing any serious attempt to modernize proving of polyphase metering connections. It appears that either the industry has neglected the subject or that the author is overconcerned.

Statistical Approaches to Selecting Domestic Meters for Test

LARRY DWON
MEMBER AIEE

J. A. MORRIS
NONMEMBER AIEE

THE AMERICAN ELECTRIC Power Company (AEP) system is an integrated and interconnected electric power network, including six principal operating companies serving parts of seven states: Michigan, Indiana, Ohio, West Virginia, Virginia, Kentucky, and Tennessee. The majority of its revenue watt-hour (whr) meters are selected for test on the basis of geographic areas, each of which include one-eighth of the meters in an operating company. Each year all meters in the geographic area are tested, with some minor exceptions, so that over a period of 8 years all meters on the system are tested once.

Meters are tested to assure accurate energy measurement. This is not only the company's responsibility, but also the concern of state public utility commissions. While the code for electricity meters[1] recommends that single-phase meters of 12-kva rating or less, which includes most domestic service whr meters, be tested at least once every 96 months, not all state commissions conform to this recommendation.

Past test results reveal that for every 100 meters tested, only four are found to be inaccurate beyond the prescribed limits. This very inefficient method of finding the meters needing attention, and the prospect that meter-testing expenses will continue to increase, motivated the reexamination of the present method of selecting whr meters for test. Because meters have been improved over the years, it was anticipated that those subject to more frequent error may be concentrated in some earlier classes. If this could be proved, then meter testing could be concentrated on those classes, and the number needing a test would be much smaller than at present.

Over the years, manufacturers have made improvements in the insulation and magnetic materials so that modern whr meters retain their calibration over much longer periods than those built previous to 1940, when stable braking magnets were introduced. Since all four meter manufacturers in this country made similar or equal changes at approximately the same time, such progressive changes now may be considered as establishing five age groups or categories, covering the period from 1910 to the present. This segregation is demonstrated in Fig. 1. Because of such improvements, present procedures require the testing of a large number of accurate meters, in order to find the small number that are in error. If selective testing were possible, less meters need be tested each year and, as inherently inaccurate models or types were uncovered, they could be retired if the economics involved justified such action. It was anticipated that a means for testing only a small statistical sample of all meters could be devised which would assure that the whole population were maintained within acceptable tolerances.

Statistical procedures are useful tools, but available literature does not deal with their application to the meter testing problem explicitly enough for easy understanding by people not familiar with statistical methods. This paper, therefore, briefly reviews available facts concerning present methods of testing meters, the more basic concepts of statistical sampling theory, and several procedures which might be applied to the meter testing problem.[2]

The foregoing implies that meters can be divided roughly into five developmental classes and the ability to maintain their accuracy is different for each class. The more modern meters are practically troublefree, as can be demonstrated by actual test data.

Statement of Problem

The meter classes on the AEP system include several types of each of four manufacturers. Types of a given vintage are believed to maintain approximately the same accuracy for any similar service environment. The problem, therefore, is to devise the most economical system of sampling, so that for each class of meter enough of them would be tested to determine with predictable precision and reliability the probable number of all meters of that class, which are within acceptable accuracy limits; and at the same time to assure that the new method is at least as reliable as the one now used and acceptable to the various state commissions. Furthermore, proper evaluation of the test data would show whether or not the test

---

Paper 59-167, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted October 24, 1958; made available for printing December 17, 1958.

LARRY DWON and J. A. MORRIS are with the American Electric Power Service Corporation, New York, N. Y.

The authors wish to thank C. L. Lucal and R. V. Keich, engineers on the AEP system, for their investigations and data with respect to some factors which influence meter errors. They also wish to thank Floyd Busch, N. P. Demos, and E. Davis of the General Electric Company.

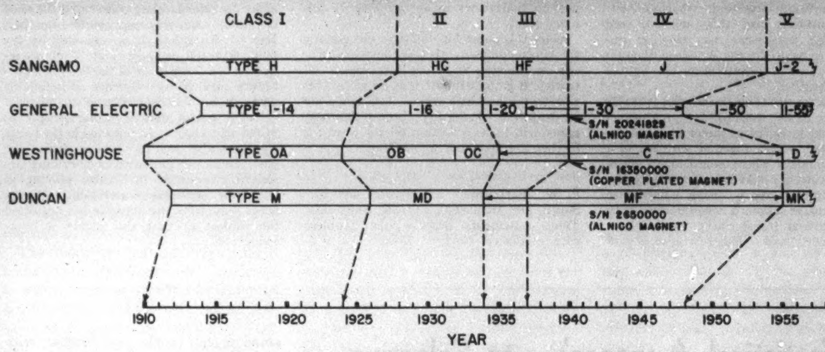


Fig. 1. Years when major domestic meter improvements occurred

data analyzed by class are truly representative of accuracy or whether further analyses by meter type is required.

The present testing method may be considered a 100% sampling of a geographic area which may not be typical of all meters on the system. Therefore, no statistical inference can be made from it about the other meters. Statistical methods may involve many kinds of sampling procedures, but all of them have certain things in common, e.g., they sacrifice accuracy for economy, but in each case the amount of accuracy is predictable. As much accuracy can be attained as one is willing to pay for.

Analyses of Present Records and Procedures

While the methods that are discussed herein apply equally well to all AEP operating companies, the specific example discussed in this paper approximates the situation for one of them. This company maintains records as follows:

1. Each meter has a history card showing the year of purchase, serial number, company meter number, tests, location, etc.
2. Each division of the company maintains these history cards segregated by manufacturer and capacity. Within a given manufacturer's group of meters, however, the various classes of meters and their location, whether urban or rural, are intermingled.
3. The company maintains a symbol book which theoretically lists the meters consecutively as to date of receipt and by various manufacturers.
4. Test slips for each meter are kept in a separate file arranged by meter number.

History cards seem to provide the best available frame of reference, from which samples of the meter universe can be selected. However, at some future time it might be advisable to refile the history cards by homogeneous study groups as shown in Fig. 2.

Meter Classes

Major improvements were made by most manufacturers at about the same time so that Fig. 1 defines approximately the types of meters in use today, segregated by class.

Fig. 3 shows the distribution of meters as follows:

1. Present rate of retirement of class I and II meters.
2. Rate of retirement that might be considered to eliminate class I, II, and III meters by 1965.
3. Rate of growth of class V and all meters on the system.
4. Rate at which class III meters are being converted to essentially the equivalent of class IV type.

Knowledge of improvements made to these meters suggests that those in class I are subject to more frequent errors and need more attention than those in class V, for instance. Special tests conducted in selected urban and rural areas substantiate the suspicion that meter errors are a function of meter class and location, as shown in Table I.

Factors Causing Meter Errors

Meter inaccuracy can be caused by several factors:

1. Age of meter: newer meters are inherently better able to maintain accuracy because they are better designed, have improved materials, and built-in protection against lightning effects.
2. Manufacturer: competition tends to keep all manufacturers' products about equally good based on years of periodic test results. This factor may merit future study, both to stimulate competitive development and improvements and to obtain benefits of the ability of one particular type to maintain accuracy better than any other type.
3. Location: local conditions, such as frequency of power surges, lightning frequency, or corrosive atmosphere, may affect meter accuracy, particularly the earlier meter types.

While other factors may also cause meter errors, the above seem to be most important. The immediate concern of this investigation, however, is the class of meter and whether its location is urban or rural.

Tests Results

The 1956 total meter accuracy for the AEP system is indicated below for the 96-month class of meter:

Distribution of Meters by Error Classes

Error, %	Per Cent Tested Light Load, 10% Rating		Full Load, 100% Rating	
0–1	66.83		86.05	
1–2	21.20	88.03	9.84	95.89
2–5	8.05		1.94	
5–10	1.34		0.44	
10 and over	1.15		0.35	
Stopped	1.43		1.38	
		100.00		100.00

This represents 221,637 meters tested over the entire AEP system.

On the basis of these data, approximately 88% and 96% of the meters tested met acceptable standards for light-load and full-load tests, respectively. Limit-

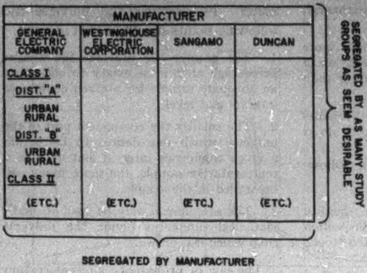


Fig. 2 (left). Meter history card file arranged by possible homogeneous study groups

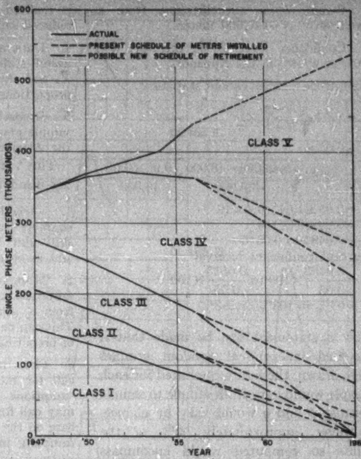


Fig. 3 (right). Meter class distribution

ing this discussion to full-load conditions, it can be said that out of each 100 meters tested, four will be outside the acceptable limits of plus or minus 2% error. It is this inefficient procedure which motivates the application of statistical sampling methods to this problem.

Statistical Samples and Inferences

The purpose of any statistical survey should be the gathering of accurate data about the object of study, i.e., the universe. One hundred per cent accuracy can be obtained only if the entire universe is included in the sample. Most often this is impracticable, very expensive, and unjustifiable.

When the universe needing investigation is relatively large, some savings, at the risk of error, can be realized by using scientific sampling methods. Scientific sampling is a particular way of selecting and investigating a part of a universe in order to find out about the whole universe. Scientific sampling is based on the mathematical theory of probability, and these concepts presuppose a known distribution of items in the universe and a similar distribution in the sample.[3]

If proper methods are used, under applicable conditions, inferences can be made about the entire universe from data obtained in relatively small samples. These inferences will lie within predictable confidence intervals, and at a known confidence level. Proper methods imply a scientific sample, random and representative, and applicable conditions imply a known distribution, preferably normal, of items in the sample and universe.

CONFIDENCE INTERVAL AND LEVEL

The following factors are inherent in the results of any sampling procedure and thus require early definition:

1. Confidence Interval. The range of precision within which the statistic of the sample will probably fall.
2. Confidence Level. The probability that this statistic of the sample will fall within that range.

These two factors are mathematically related in the Appendix. The narrower the range, the less probable that a given sample will fall within the range. This is turn depends upon the type of distribution of items in the sample and universe and the statistic being measured.

The confidence interval is expressed mathematically in equations 1 and 3. The following example may clarify the interpretation of the foregoing terms:

1. Sample items = 48,000 meters actually tested.
2. Proportion with less than ±2% inaccuracy $= p = 0.95$.
3. Proportion with more than ±2% inaccuracy $= q = 1 - p = 0.05$.

Table I. Meter Accuracy Related to Date of Manufacture, July 1957 Tests

	Division 1, District C					Division 2, District α					Division 3, District VW				
Class	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
Date of Manufacture	1914 to 1926	1926 to 1934	1934 to 1940	1940 to 1954	1950 on	1914 to 1926	1926 to 1934	1934 to 1940	1940 to 1954	1950 on	1914 to 1926	1926 to 1934	1934 to 1940	1940 to 1954	1950 on
Urban															
More than 2% fast	13.4	11.6	10.4	0.5	0	4.5	1.4	3.4	0	0	3.7	1.2	1.2	0.5	0
More than 2% slow	1.2	2.9	0	0.2	0	0	2.9	1.1	0	0	0.4	4.9	0.4	0.2	0
More than 2% error	14.6	14.5	10.4	0.7	0	4.5	4.3	4.5	0	0	4.1	6.1	1.6	0.7	0
Number in sample	253	69	183	577	147	222	69	89	201	5	268	81	243	400	74
% of number of meters tested	20.6	5.6	14.9	46.9	12.0	37.9	11.8	15.2	34.2	0.9	25.2	7.6	22.8	37.5	6.9
Total number of tests			1,229					586					1,066		
Rural															
More than 2% fast	47.4	0	12.9	0	0	22.5	18.5	19.5	1.3	0	29.8	25.0	33.8	2.1	0
More than 2% slow	0	0	0	0	0	0	0	1.2	0.6	0	1.1	0	0	0	0
More than 2% error	47.4	0	12.9	0	0	22.5	18.5	20.7	1.9	0	30.9	25.0	33.8	2.1	0
Number in sample	19	2	31	99	51	169	27	164	527	22	94	28	65	142	28
% of number of meters tested	9.4	1.0	15.3	49.0	25.3	18.6	3.0	18.0	58.0	2.4	26.3	7.8	18.2	30.9	7.8
Total number of tests			202					909					357		

4. Sample is a scientific one from a universe of $N=440{,}000$ meters.

5. Confidence level = 90%, therefore $k=1.64$, as shown in the Appendix.

Confidence interval (about statistic q)

$$=q\pm k\sqrt{\frac{N-n_1}{N-1}\times\frac{pq}{n_1}}$$

$$=0.05\pm 1.64\sqrt{\frac{440{,}000-48{,}000}{439{,}999}\times\frac{0.95\times 0.05}{48{,}000}}$$

$$=0.05\pm 1.64\times 9.4\times 10^{-4}$$

$$=0.05\pm 0.00154$$

$$=0.0485 \text{ to } 0.0515$$

Per-cent confidence interval

$$=\frac{0.0485}{0.05}\times 100 \text{ to } \frac{0.0515}{0.05}\times 100$$

$$=97\% \text{ to } 103\% \text{ or } \pm 3\%$$

Now a statement can be made that if repeated independent random samples were drawn, the limits computed for each sample would vary from sample to sample because p and q would vary by chance. However, approximately 90% of the limits so computed would encompass the true proportion q of the universe.

We would make a similar interpretation for other statistics that might be used such as the arithmetic means $\bar{X}_s$ in the Appendix.

STATISTICAL APPROACHES

Two statistical approaches have been used extensively in surveys, quality control, and other work, namely, method of Attributes and method of Variables. The Attributes method provides a means for making a statistical inference from a sample on the basis of the statistics p and q, where

p = proportion of sample meters within test limits
q = proportion of sample meters outside test limits and
$p+q=1$

The method of Variables provides a means for making a statistical inference from a sample on the basis of the statistics $\bar{X}_s$ and $\hat{\sigma}_u$, where

$\bar{X}_s$ = arithmetic mean of sample meter test accuracy
$\hat{\sigma}_u$ = estimate of standard deviation of universe

The Variables method has the following advantages as compared to the Attributes method:[4]

1. For any desired degree of precision and reliability, fewer items are needed in the sample for single sampling procedures.
2. Measurement of a sampling item under the Variables method gives much more information about that item than under the Attributes method.
3. The Variables method will yield more information about the universe than would the Attributes method. This advantage is realized in a practicable way through reduced sample sizes.
4. Data are made available about more useful statistics, namely arithmetic mean $\bar{X}_s$ and standard deviation $\hat{\sigma}_u$ rather than proportions p or q.
5. Borderline items in the Variables sample present no problem, but they do in the Attributes method.

The Variables method has the following disadvantages:

1. Variables inspection of a quality characteristic requires more inspection skill and time, more complicated tools, and in other ways is more costly.
2. More record keeping and more arithmetic are needed after the data are obtained from a sample in the Variables method. This can be minimized by more initial work of preparing graphical charts and tables.
3. Plans based on the Variables method call for more stringent mathematical assumptions. Their selection and installation may call for considerably more ability and care on the part of those administering the sampling inspection program.

As far as the need for more complicated testing facilities mentioned, the disadvantage in not especially pertinent since meters have to be tested. The relative merits appear to favor the Variables method.

UNIVERSE AND ITS DISTRIBUTION

In statistics the universe means all items that comprise the object of study. It is sometimes referred to as the population. In this case it is about 440,000 domestic meters which are being tested on a 96-month cycle. This universe can be limited, if desired to some homogeneous part. For instance, we might want to study only the meters in some district, division, or geographic area. Perhaps we may want only to study one age class or one manufacturer's type of meter. In selecting the universe the following points should be kept in mind:

1. The greater the universe the smaller percentage of it is necessary to constitute an adequate sample for a given confidence interval and level.
2. The smaller the component part of the universe which one desires to be within a given confidence interval and level in a representative sample, the more items will be needed in the sample.
3. The sample can only be used to make statistical inferences about the universe from which it is drawn.

The law of probability suggest certain patterns or distributions into which chance events seem to fall. Chief among these are the Binomial distribution and its approximations, the Poisson distribution, and the Gaussian or Normal distribution. Our problem requires the use of probability concepts, but we will presuppose in our discussion, that if the Attributes method is used, enough items are included in the sample to allow the use of simpler concepts of the Normal distribution. When the sample is large enough, the Normal theory is applicable and these conditions prevail when the following conditions are satisfied:[5]

Observed Proportion p	Sample Size for Normal Theory to Apply
0.4 or 0.6	50
0.3 or 0.7	100
0.2 or 0.8	400
0.1 or 0.9	1,000

Actually our p values are in the order of 0.96 which means that to apply Normal distribution concepts we would penalize ourselves with an unnecessarily large sample. If Normal theory does not apply, charts or tables prepared directly from

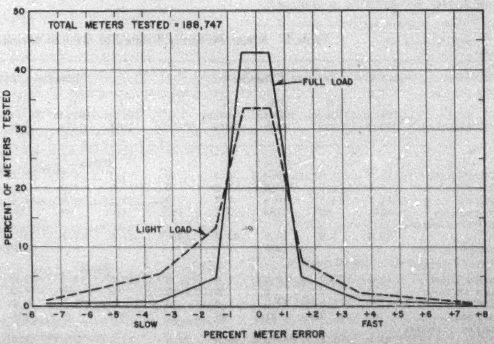
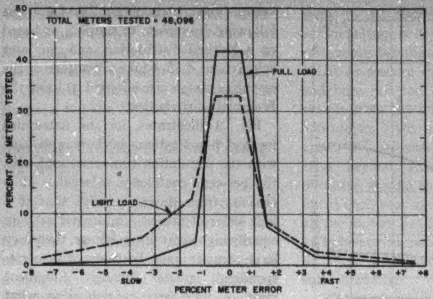


Fig. 4. AEP system meter accuracy frequency distribution, 1956



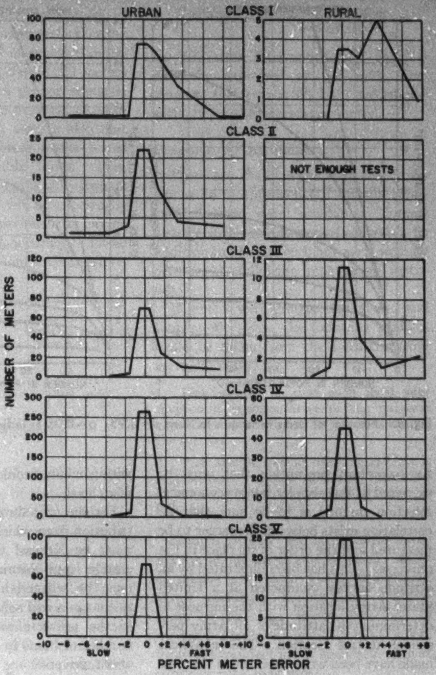


Fig. 5 (above). Subsidiary company system meter accuracy frequency distribution, 1956

Fig. 6 (right). Subsidiary company division meter accuracy frequency distribution, July 1957 tests

the Binomial distribution are available.[6]

When applying the Variables method it is required that the test samples and the universe be Normally distributed. These should be tested by available statistical methods. Therefore, one should obtain a past history of frequency distributions of meter errors early in the application of statistical approach.

In spite of the insistence by some that the Normal universe is not natural or universal, a great many experimental, manufacturing, economic, and other processes do produce distributions that are fairly near Normal. Figs. 4 through 6 illustrate this contention in varying degrees for the following situations:

Fig. 4: AEP system total meter accuracy frequency distribution (full and light load) in 1956.

Fig. 5: Subsidiary company total meter accuracy frequency distribution (full and light load) in 1956.

Fig. 6: Subsidiary company special tests in division 1 on classes I to V meters in rural and urban areas.

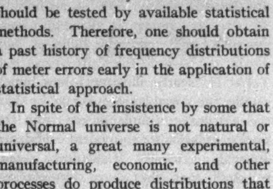


Fig. 7. Area between selected ordinates of normal curve

The Normal distribution is symmetrical and unimodel. Hence, its mean, median, and mode are all the same. The fractions of the Normal universe lie within certain symmetric limits as shown in Fig. 7.

Scientific Sample

A scientific sample is selected at random and has the following characteristics:

1. Each item is drawn independently.
2. Each item has the same probability of being drawn as any other one.
3. The homogeneous items in the sample are proportionately distributed, or representative, in the same manner as in the universe.
4. Such a composite of items reflects the universe from which it was drawn, within a predictable error of estimate and a known confidence level.

A truly random sample is sometimes difficult to obtain because the frame of items in the universe may not be available or is in such a form as to make it impracticable to use properly. The simplest way to visualized the selection of a random sample is to use a drum in which the universe is represented by consecutively numbered small marbles. The container contents are continuously mixed and items are withdrawn blindly one at a time. This procedure is obviously cumbersome.

One method, called "simple random" selection, which accomplishes the objective with more finesse, is to utilize random number tables. Each element in the universe is assigned a number. These numbers may be the place order of meter history cards. Successive numbers of the appropriate size are read from the table, repeated numbers being discarded, until the required sample size is attained. The elements of the universe with corresponding numbers constitute the sample.

Another method, called "systematic random" selection, can be used when applicable. In this case one enters the random number table to select blindly a single numbr which falls within the group of numbers of the universe. Then every ith number is chosen as part of the sample where i equals the following:

$$i = \frac{\text{Items in universe}}{\text{Items in sample}}$$

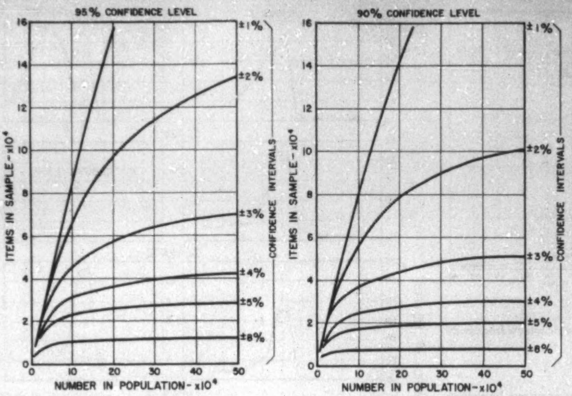


Fig. 8. Number of items in sample for case $p=0.95$, $q=0.05$ at indicated confidence levels

An important precaution that must be exercised in this simpler, more economical selection method is to be sure that no correlation exists between the factor to be measured and the order of its place in the universe. This has been illustrated by an attitude survey conducted at a United States Army location with the purpose of determining the attitude of all Army personnel at this location. The inferences made have been wrong because in this case every *i*th item of the universe happened to be mostly sergeants. The sample therefore was biased by opinions of sergeants.

STATISTICAL INFERENCE

The process by which, on the basis of a sample, a conclusion is made about the universe from which the sample is taken, is called "inference." A statistical inference is always a conclusion about the value of a statistical measure describing the universe: its mean, standard deviation, proportion, etc.

In this study we will consider inferences about the proportion of domestic meters which exceed ±2% error (method of Attributes). Thus, on the basis of the random sample we will judge what proportion of meters in the universe exceeds ±2% error and within what "confidence interval" and on what "confidence level."

We will also consider inferences about the arithmetic mean $\bar{X}_s$ and estimated standard deviation of the universe $\hat{\sigma}_u$ obtained from the meter accuracy test samples (method of Variables). Thus on the basis of the random sample we will judge the arithmetic mean and standard deviation of the meters in the universe and within what confidence interval and on what level.

For the procedure to be valid, the distribution from which the sample is taken must be Normal or depart only moderately from Normal shape; the sample must be selected in a scientific manner so that in a real sense it is an "offspring" of the universe as a whole; and the number of items in the sample (evidence about universe) are sufficient to make the error of estimate reasonable.

ITEMS IN SAMPLE

The number of items in the sample picked from a normally distributed universe is given by equation 2 and 4. The following discussion, to simplify, is based on a Normal distribution, and our later calculations do not take advantage of the saving in sample size afforded by using the Binomial distribution.

Fig. 8 illustrates, for the Attributes method, how the items in the sample vary with the population (universe) and with the per-cent confidence intervals. It is obvious from these diagrams that if a high order of precision is desired with an equally high order of confidence, then very large samples must be taken with correspondingly greater expense. Judgment must be used here to determine what limits to select and with what confidence to be satisfied. These will be rationalized in a following discussion.

Typical values of $\hat{\sigma}_u^2$ and corresponding values of pq from the same sampling distribution are compared in Table II. These values were calculated from test results shown in Figs. 4 and 6.

Table II shows the ratio of n_2/n_1 for the different combinations of $\hat{\sigma}_u^2$ and pq values already mentioned. This ratio was calculated by the following equation derived from equations 2 and 4 where the symbols are defined:

$$\frac{n_2}{n_1}=\frac{\hat{\sigma}u^2}{pq}\times\left(\frac{w^2(N-1)+k^2pq}{w^2(N-1)+k^2\hat{\sigma}_u^2}\right)$$

Table II clearly illustrates how much smaller the Variables sample may be than the Attributes sample picked from the same population for equivalent confidence intervals and levels. It should be noted, however, that the above formula exaggerates the n_2/n_1 ratio because the best

Table II. Ratio of Variable to Attribute Sample Sizes for Equal Confidence Level and Interval and Universe

		Variances (Multiply by 0.0001)					
		Variables, $(\sigma\mu)^2$		Attributes, pq		n_2/n_1*	
Meter Class	Operating Division	Urban	Rural	Urban	Rural	Urban	Rural
I	1	5.25	5.04	1275	2440	0.006	0.004
	2	2.38	5.75	471	1745	0.006	0.004
	3	3.03	7.29	462	2170	0.008	0.006
II	1	7.64		1240		0.009	
	2	1.84	8.49	414	1730	0.005	0.008
	3	3.81	6.03	579	1875	0.008	0.005
III	1	4.94	6.08	975	1120	0.007	0.008
	2	2.76	7.05	430	1640	0.007	0.007
	3	1.68	6.68	162	2224	0.011	0.006
IV	1	0.79		69		0.011	
	2		1.85		204		0.010
	3	1.06	2.10	75	274	0.015	0.009
All	AEP	2.19		394		0.007	
All	Subsidiary companies	2.74		357		0.009	

* These values were calculated with the following conditions:
$N=50{,}001$
$w=0.08q=0.08\times0.05=0.004$
$k=1.64$ (90% confidence level)
Range of $n_2/n_1=0.004-0.015$
Average of $n_2/n_1=0.008$

Attributes sample is not utilized, as explained previously. To design the best Attributes plan the concepts used in quality control, the control charts, should be utilized. These are based on the Binomial not the Normal distributions.[7,8]

Applying Statistics to Meter Testing Problem

We will assume that the universe and samples drawn from it are Normally distributed. To increase this possibility under the Attributes method, no sample of a subgroup in which we are interested will contain less than 1,000 items unless it is physically impossible to do so. Under the Variables method no sample of a subgroup will contain less than 30 items.

While we have no experience to say what is a reasonable confidence interval and level, it might be possible to make an estimate from various aspects of our present procedures. Present practice obtains almost a 90% sample, approximately 48,000 meters out of 55,000 (1/8 of 440,000), from a geographic area and we actually realize almost a zero confidence interval on a 100% confidence level because practically all meters are tested in that specific universe. Those meters that exceed the limit of error are recalibrated or replaced.

However, we cannot make any statistical inference about the meters in other geographic areas until their test period arrives once each 8 years. Practical judgment leads us to believe that the more recently tested meters will have a smaller percentage in error than those which have not been tested for 7 years, for instance. This means that, at the end of a year, approximately 5,500 meters would be expected to remain in error over the system as follows:

Meters	Meters Assumed in Error of 55,000 Each Year: Estimated %	Meters Assumed in Error of 55,000 Each Year: Meters
Recently tested	0	0
Years without test		
7	4	2,200
6	3	1,650
5	2	1,100
4	1	550
3	0	0
2	0	0
1	0	0
Total		5,500

These 5,500 meters may be assumed in error on the system plus some unknown quantity, possibly 4% of 7,000 meters (55,000 to 48,000), which were not tested in the geographic area. This would seem to be inadequate confidence in our knowledge of how many meters on the system are beyond the inaccuracy limits.

Let us assume that the 48,000 meters had been selected as a proper scientific sample from the whole universe of 440,000 meters. We must assume also that about 2% of the meters might be in error beyond acceptable limits. This, of course, is what the sample will estimate for us each year and it will be affected by all the factors listed previously as having an influence on meter inaccuracy. Based on these assumptions our calculated results would be:

Per-Cent Confidence: Level	Per-Cent Confidence: Interval	Probable Range of Meters in Error
99	±4.7	8,800† ±413
95	±3.5	8,800 ±308
90	±3.0*	8,800 ±264

* See previous illustrative computation.
† 0.02×440,000 = 8,800.

Thus with present procedures we estimated a possible range of meters in error of 5,780 (5,500+0.04×7,000) which is many times the ranges computed above. Furthermore, however, we do not have now any idea of how confident we can be that the meters beyond ±2% inaccuracy equal zero plus or minus 2,890 (5,780/2). Under sample testing we are able to compute the probable range of meters in error and what the probability is that this range encompasses the true number of meters in the universe which exceed the prescribed accuracy limits.

At best we can say that our present method leaves us with a fairly wide range of doubt, 5,780 meters, and with no predictable confidence level. A ±8% interval would have a range of 1,406 meters (8,800×.08×2) at a confidence level of 90%, over 400% better than present practice. We also know that a level of 99% and a narrow interval like ±0.1% are used mostly in cases involving life and death decisions. This is not such a case. In quality control a 95% level is used very often.

The subsequent discussion of the costs involved in various possible statistical approaches will give us additional help to judge what confidence level and interval to use. We may now, however, suggest that a 90% level ±8% interval would seem to be reasonable for this problem as a start and that the interval can probably be increased.

It should be noted that there are two ways of interpreting the confidence interval percentage, ±8%, as follows:

1. Confidence interval $= \bar{X}_s \pm 0.08\,\bar{X}_s$ or $= p \pm 0.08\,p$
2. Confidence interval $= \bar{X}_s \pm 0.08$ or $= p \pm 0.08$

In the latter interpretation $\bar{X}_s$ and p must be expressed as per unit. Our usage of the term ±8% is in accordance with the first interpretation noted.

POSSIBLE APPROACHES

There are many ways that this problem could be approached if suitable reference frames were available. Present history cards provide a practical way. In addition to this method, however, we may consider the following hypothetical methods.

If meters were grouped by classes and local subdivisions, the following samples could be taken from the total company meter universe:

1. Pick sample so that each local subdivision is representative of the universe regardless of which meter class falls into the district.
2. Pick sample so that each meter class is representative of the universe, regardless of which district or geographic area it falls into.

Sample 1 would yield some useful information about variations between classes of meters and more useful information about variations among local subdivisions or, in other words, locations of meters. The latter would tell us where local factors affect meter accuracy, such as urban or rural exposure.

Sample 2 would provide a means of making an inference about all meters in the system universe and about individual classes as well. If we assume that our

Table III. Number of Meters in Hypothetical Samples

Attribute Method: Obtained From All Domestic Meters in a Universe of 440,000

Class of Meter	Number of Meters: Class	Samples: A	Samples: B	Samples: C
I	83,750	1,275	2,970	7,350
II	33,750	1,000*	1,400	6,500
III	57,500	1,000*	2,270	7,100
IV	187,500	3,230	7,500‡	7,500
V	77,500	1,425	3,320	7,300
Total	440,000	7,930 †	17,460	35,750§

Case A: Sample selected so confidence (90%; ±8%) applies to whole universe only.
Case B: Sample selected so confidence (90%; ±8%) applies to class IV only.
Case C: Sample selected so confidence (90%; ±8%) applies to all classes equally well.
* Minimum number for normal distribution to apply.
† Only 7,500 items would be required if foregoing condition were disregarded.
‡ Only class IV meets confidence; all others are poorer.
§ Not a representative sample of the whole universe.

Table IV. Number of Meters in Practical Samples

Attribute Method: Obtained from All Domestic Meters in a Universe of 440,000

Division Name	Number of Meters			
		Samples		
	Division	D	E	F
Southern	146,500	2,490	7,000*	7,000
Western	117,000	2,000	5,610	7,000
Central	111,200	1,910	5,350	7,000
Eastern	65,300	1,100	3,120	6,900
Total	440,000	7,500	21,080	27,900†

Case D: Sample selected so confidence (90%; ±8%) applies to whole universe only.
Case E: Sample selected so confidence (90%; ±8%) applies to southern division only.
Case F: Sample selected so confidence (90%; ±8%) applies to all divisions equally well.
* Only southern division sample meets confidence; all others are poorer.
† Not a representative sample of the whole universe.

confidence applies only to the system as a whole, then the result would be case A, the smallest sample for the conditions. For any individual class however, our confidence would be less but still a predictable one. If we desire inferences about the largest individual class to be at the established confidence level, case B, a larger number would be required in the sample; the confidence level for the whole group would be greater, and the inferences about the smaller classes would be on lower confidence levels but still predictable ones. If on the other hand we desire the smallest class to have the reliability and precision, then inferences about all other classes including the universe would have a higher reliability and precision and this would require a very much larger sample. This case is disregarded because it is impracticable.

We might be interested in another situation, case C, as shown in Table III, in which a sample would be selected that is not proportionately representative of the universe, but which does insure each class inference to have the same precision and reliability.

Practical Approach

This company's meter history card file intermingles classes and locations within each manufacturer's group. Since we are not interested at this time in determining differences among manufacturer's products, a systematic sample can be picked from the total company meter universe so that each division is representative in the sample.

The meter classes and their locations are assumed to be intermingled in a random manner. It is reasonable to expect therefore, that a sampling distribution would have the same proportion of meter classes as the system. It is also possible that the sampling distribution would have the same proportion of urban and rural meters as the system. This, of course, should be checked when samples are drawn because concentrated rural programs may have caused meter history cards to be grouped in a manner other than random.

The sample items necessary to satisfy Attributes sampling cases A through C are given in Table III. In Table IV, similar data are given for cases D through F based on the practical conditions mentioned above. Table V shows the sample items that would be needed under the Variables method for case G, which corresponds essentially to cases A and D, and case H, which corresponds to cases C and F. These data are based on a n_2/n_1 ratio of 0.1 and the requirement that no sample shall be less than 30 in any study group. This condition insures slightly better precision and reliability than would be obtained from equivalent Attributes samples.

Table V. Number of Meters in Practical Samples

Variable Method: Obtained from All Domestic Meters in a Universe of 440,000

Division Name	Number of Meters		
		Samples*	
	Division	G	H
Southern	146,500	67	70
Western	117,000	54	70
Central	111,200	52	70
Eastern	65,300	30	70
Total	440,000	203	280†

Case G: Sample selected so confidence applies to whole universe only.
Case H: Sample selected so confidence applies to all divisions equally well.
* Samples selected on basis of $n_2 = 0.01\ n_1$.
† Not a representative sample of the whole universe.

Cost Comparisons

For this study we assume that it costs approximately $1.49 to test a meter.

Table VI shows the cost necessary to test the number of meters that cases A through H require, as well as the approximate difference in testing cost when compared to the present method. This table compares differences on the 99%, 95%, and 90% confidence levels each with a ±8% confidence interval.

We cannot say that these differences in testing cost really represent immediate savings because no reduction in meter-department expense can be effected until the equipment, personnel, and supervision cost of the meter department is reduced. Actually, it is proposed that initially as many meters be tested as necessary but the tests should concentrate on those meters which tend to develop errors sooner. As meter types are uncovered,

Table VI. Cost Comparison of Several Meter Testing Samples

Estimated Difference of Statistical Sampling Methods Over Present Procedure. All Testing Costs Assumed to be $1.49 Per Meter

Universe from Which Samples Were Selected	Cases Being Compared	99% Confidence Level and ±8% Confidence Interval			95% Confidence Level and ±8% Confidence Interval			90% Confidence Level and ±8% Confidence Interval		
			Dollars			Number			Dollars	
		Number of Meters	Test Cost	Difference	Number of Meters	Test Cost	Difference	Number of Meters	Test Cost	Difference
Geographic area	Present method	48,000	71,500		48,000	71,500		48,000	71,500	
Entire Company	A*	19,000	28,300	43,200	11,590	17,250	54,250	7,930	11,800	59,700
	B*	40,730	60,700	10,800	25,570	38,000	33,500	17,460	26,000	45,500
	C*	76,600	114,000	(42,500)	48,950	73,000	(1,500)	35,750	53,200	18,300
	D†	19,000	28,300	43,200	11,500	17,250	54,250	7,500	11,200	60,300
	E†	48,140	71,700	(200)	31,610	47,200	24,300	21,080	31,400	40,100
	F†	66,250	98,750	(27,250)	41,500	61,800	9,700	27,900	41,600	29,900
	G‡							203	302	71,200
	H‡							280	417	71,080

* Hypothetical cases using samples illustrated in Table III, Attribute method.
† Practical cases using samples illustrated in Table IV, Attribute method.
‡ Practical case similar to cases E and F, Variables method.

Table VII. Comparison of Incremental Decreases in Testing Costs for Corresponding Incremental Increases in Confidence Level

Statistical Sampling Methods, Cases A Through F

		Per-Cent Change for Indicated Changes in Confidence Level					
		90% to 95%		95% to 99%		90% to 99%	
Universe from which Samples Were Selected	Cases Being Compared	Decrease in Cost Differential	Increase in Confidence Level	Decrease in Cost Differential	Increase in Confidence Level	Decrease in Cost Differential	Increase in Confidence Level
Entire company	A*	9.0	5.5	20.0	4.2	27.0	10
	B*	17.0	5.5	68.0	4.2	76.0	10
	C*	‡	5.5	‡	4.2	‡	10
	D†	10.0	5.5	20.0	4.2	38.0	10
	E†	39.0	5.5	‡	4.2	‡	10
	F†	68.0	5.5	‡	4.2	‡	10

* Hypothetical cases using samples illustrated in Table III, Attribute method.
† Practical cases using samples illustrated in Table IV, Attribute method.
‡ Procedure costs more than present methods.

by analysis of test data which seem to be inherently subject to frequent error, they can be eliminated gradually through retirement, if testing costs compared to replacement costs are unfavorable; or they may be tested at more frequent intervals than once in 8 years. The procedure that should be followed will necessarily have to be developed by experienced meter engineers.

The company already knows which class of meters is the more inaccurate but these data are based on a very small part of its meter universe. However, it is obvious that while the incorrect meter classes are known, one cannot tell which of the meters in those classes are in error. For instance, if it were assumed that meters in classes IV and V would maintain calibration indefinitely, an unreasonable assumption, the company would still need to test about 175,000 meters, 40% of total, this year to locate approximately 5,600 inaccurate meters, 1.27% of total. This, of course, cannot be done in a single year because with present facilities and personnel, only 12$^1/_2$% are tested per year therefore consideration must be given to the following:

1. The elimination, as it becomes practicable, of those meter classes which contain inherently inaccurate meters.
2. The utilization of different testing cycles for the various classes of meters.

Eventually, as the meter universe becomes composed of classes IV and V meters and perhaps others containing even more accurate types, it may well be that a smaller sample need be tested than at present. At that time, it appears that only between 4,000 and 8,000 meters need be tested, if the Attributes method were used, and perhaps something like 200 to 300 meters if the Variables method were used, to assure a degree of reliability and precision now believed to be reasonable. This is based only on study groups mentioned in this report. At this future time, possibly 1965, an appreciable reduction in testing costs may be expected, and what is more important, the cost will not increase greatly as more domestic meters are added.

Returning to Table VI and the foregoing discussion, it now seems possible to show that a 90% confidence level with a minimum of ±8% confidence interval is more than adequate for our problem at present if the Attributes method is used. Table VII compares the incremental increases in the confidence level and the corresponding decreases in differential costs that occur. Taking class A as an example, Table VII shows that if we increase the confidence level from 90% to 95%, an incremental increase of 5.5%, we would decrease the differential cost by 9%, (a faster rate). If we were inclined to go from 90% to 99% the incremental gain in confidence would be 10%, but with an incremental decrease in differential cost of 27%. Other cases are even more extreme. Is it prudent, therefore, to require the increased reliability for this problem? The authors think that it is not, but this needs to be thoroughly reviewed and discussed so that reasonable and acceptable limits are established consistent with all ramifications of the problem.

Conclusions

The following major conclusions result from this study.

PRESENT METHODS OF TEST

1. Meter-testing expense has been increasing each year and will continue to increase as additional customers are added, unless different procedures are devised.

2. Present selection methods are inefficient. Of each 100 meters tested only 4 are found to be in error.

3. Evidence is available which shows that different classes of meters have different frequency of error. These data presently are limited, localized, and inapplicable for statistical inferences about all meters of a system.

4. Evidence is also available which shows that rural meters show higher frequency of error than urban meters. These data also are limited, localized, and inapplicable for general statistical inferences.

5. Present test procedures tell us exactly how many meters are in error and to what degree in a geographic area only. They tell us nothing about the other 87.5% of the meters on the system, most of which will not be tested for seven years.

STATISTICAL METHODS

1. Statistical sampling methods provide a means for making inferences about a large number of items, like meters, on the basis of relatively small samples. These inferences can be as precise and reliable as one wishes them to be. The precision and reliability is mainly determined by the number of items used in the sample and this, in turn, affects the expense of the survey.

2. By judicious selection of desired precision and reliability, savings can be realized in seeking information about the meters with which we are concerned.

3. Until further information proves otherwise, it is suggested that we use a 90% confidence level and ±8% confidence interval as a minimum degree of reliability and precision respectively. There is a reason to believe that practical considerations might dictate a different precision.

4. The advantages of the Variables method surpass the disadvantages so its use is recommended over the Attributes method. The principal practical advantage is the smaller sample needed for a given precision and reliability. This conclusion is based on single sampling methods only.

APPLICATION OF STATISTICS TO THE SELECTION OF METERS FOR TEST

1. Statistical sampling methods are applicable to this problem; first to obtain information about the relative errors in different study groups (meter classes, location, etc.) and second, to sample from the whole meter universe in an operating company to determine significant changes in meter accuracy from year to year.

2. It is estimated that by utilizing

statistical sampling the number of meters needing test on a system with 440,000 meters might be reduced from 48,000 to about 500 some time in the future when and if all inherently inaccurate meters are eliminated. In the interim period a sample of 1,000 meters is suggested to obtain sufficient data about the universe, e.g., the relative accuracy of different classes of meters and how the accuracy is related to location and other factors.

3. If suitable changes are made in the meter department to take advantage of anticipated reduction in testing needs, then considerable savings in testing costs might be realized after all inherently inaccurate domestic meter classes are eliminated. This presupposes that a situation might some day prevail when the average error of the system meters will be so small as to be insignificant. Of course, if a significant change should occur, then testing in addition to the sample will be required which would decrease the savings that would be realized.

Nomenclature

N = number of items in universe

n_1, n_2 = number of items in sample

p = proportion of items in sample possessing characteristic being measured in decimal form or

$$= \frac{\text{number of meters within } \pm 2\% \text{ error}}{n_1}$$

$q = 1 - p$

w = one half of total desired width at a specific confidence interval in decimal form

k = 1.64 for 90% confidence level
1.96 for 95% confidence level
2.58 for 99% confidence level

$\frac{N-n_1}{N-1}, \frac{N-n_2}{N-1}$ = finite correction factor

$\bar{X}_s$ = arithmetic mean of sample

$\hat{\sigma}_u$ = estimate of standard deviation of universe from sample of n_2 items

$$= \sqrt{\frac{\Sigma_z{}^2}{n_2 - 1}}$$

Appendix. Some Statistical Equations Applicable to this Problem, Method of Attributes

The following assumptions will be made:

1. A scientific sample has been picked with no bias caused by method of selection.

2. Universe and sample are Normally distributed.

Other methods utilizing the Binomial distribution yield smaller Attributes samples than this one, which assumes a Normal distribution.

Confidence interval =

$$p \pm k\sqrt{\frac{N-n_1}{N-1} \times \frac{pq}{n_1}} \qquad (1)$$

n_1 = estimate of required sample size =

$$\frac{k^2pq}{w^2\frac{(N-1)}{(N)} + \frac{k^2pq}{N}} \qquad (2)$$

Confidence interval =

$$\bar{X}_s \pm k\sqrt{\frac{N-n_2}{N-1} \times \frac{\hat{\sigma}_\mu}{\sqrt{n_2}}} \qquad (3)$$

n_2 = estimate of required sample size =

$$\frac{k^2\hat{\sigma}_\mu{}^2}{w^2\left(\frac{N-1}{N}\right) + \frac{k^2\hat{\sigma}_\mu{}^2}{N}} \qquad (4)$$

References

1. CODE FOR ELECTRICITY METERS. *C12-1941*, American Standards Association, New York, N. Y., fourth edition, 1941.

2. SAMPLING METER-TESTING TECHNIQUES, **W. Capers Harrison.** *Public Utility Fortnightly*, Public Utilities Reports, Inc., Washington, D. C., vol. 59, May 9, 1957.

3. METHODS OF STATISTICAL ANALYSIS IN ECONOMICS AND BUSINESS (book), **E. E. Lewis.** Houghton Mifflin Company, Boston, Mass., 1953, pp. 17–192.

4. SAMPLING INSPECTION BY VARIABLES, **A. H. Bowker, H. P. Goode.** McGraw-Hill Book Company, Inc., New York, N. Y., 1952, p. 4.

5. RESEARCH METHODS IN SOCIAL RELATIONS, PART II (book), **Marie Johada, Morton Deutsch, S. W. Cook.** The Dryden Press, New York, N. Y., 1953, p. 652.

6. TABLES FOR STATISTICIANS (book), **H. Arkin, R. R. Colton.** Barnes & Noble, Inc., New York, N. Y., 1950, p. 134.

7. STATISTICAL QUALITY CONTROL, **E. L. Grant.** McGraw-Hill Book Company, Inc., second edition, 1952.

8. STATISTICAL QUALITY CONTROL AND ACCEPTANCE SAMPLING. *ORD-M-608-9*, U. S. Government Printing Office, Washington, D. C.

◆

Discussion

F. H. Busch (General Electric Company, Rochester, N. H.): Mr. Dwon and Mr. Morris are to be complimented on a fine paper which will add materially to the rapidly growing body of knowledge covering the application of statistical techniques to the measurement of whr meter performance. These techniques have been developed and researched in great detail by theoretical mathematicians and statisticians. Quality control engineers are working diligently and successfully to reduce this technical knowledge to methods and plans that can be applied to everyday industrial problems. As the methods are applied to a new field, a considerable number of people must learn enough of the fundamentals and principles used in order that proper applications can be made. This paper serves this purpose by illustrating and discussing the use of statistical methods for the problem of measuring the performance of whr meters. Operating meter superintendents will find this paper useful and helpful to the consideration of their own situations.

It is to be expected that various individuals will propose and defend a variety of other procedures for selecting the sample, determining the number of meters in the sample, and classifying or grouping the meters for evaluation. The point to remember is that the number of possible specific sampling plans or procedures is infinitely large though all are based upon a few fundamental mathematical concepts. The requirements and objectives of the particular application will determine the particular details of the best plan for that situation.

The authors have used the concepts of confidence level and confidence interval as a measure of the reliability and precision of the sampling plan to be used. These are entirely valid means for such evaluation, though I have found it difficult for persons who have not studied probability theory to relate these measurements to the needs of a practical application. An alternative approach that I have found useful makes use of probability statements such as the following:

"The quality of meter performance is adequate when the results of the sample-lot evaluation provide assurance such that there is only one chance in ten that the per-cent of accurate meters is less than A%. A may be 98%, 97$^1/_2$%, 97% or any per cent which the operating superintendent considers to be the poorest level of performance that he can accept without taking specific and above normal corrective actions.

This approach results in a larger sample size when the actual level of performance is near A% and a smaller sample size as the actual level is increasingly better than A%.

Since actual results of previous tests provide an evaluation of the probable true per cent of meters in error, it is possible to determine the sample size required for the ensuing year with consideration of the particular situation then existing. This approach does require that the operating superintendent determine the level of metering performance which is adequate and acceptable in his situation. It also assumes that separate, specific, and effective corrective actions will be taken as the actual level of quality is found to be approaching the minimum acceptable.

Many factors must be evaluated in arriving at this minimum acceptable quality level, including technical feasibility, costs of attaining improvement, revenue losses, customer complaints and relations problems, and evaluation of regulatory requirements. However, these are evaluations which operating meter superintendents have been making as a routine part of their jobs for many years. The only new factor is the need to reduce those evaluations to the simple probability statement with a specific minimum percentage of accurate meters being defined and established. Once this is accomplished the particular sampling plan and a program of administra-

Vol
Pt.
195

tion can be worked out with the use of the concepts presented in the paper or with the aid of a competent statistician or quality control engineer.

As a final comment on this paper, I am pleased that the authors recognized that statistical approaches to the selection of domestic meters for test do not result in cost savings until the number of meters in error is insignificant or, stated differently, the accuracy performance is better than minimum acceptable level.

My own experience with the application and use of these quality control techniques in the manufacture and factory calibration of whr meters has been that both better quality and lower cost are realized when they are used as compared to when they are not. It is because of this experience and the conviction that similar benefits, better accuracy in service with lower testing costs, can be achieved by the meter departments of electrical utilities that I commend Mr. Dwon and Mr. Morris for providing this paper for the consideration of their fellow meter engineers.

J. G. C. Templeton (Hydro-Electric Power Commission of Ontario, Toronto, Ont., Canada): The authors are to be commended for their initiative in proposing stratified random sampling as a method of selecting domestic meters for test, and for working out in some detail alternative sampling plans for their company's population of meters. The proposal to group the many types of meters in the population into five classes for sampling purposes is particularly useful. A number of statements in the paper are open to criticism from the point of view of theoretical statistics. Since the paper is not written for or by professional statisticians, however, it would be inappropriate to pounce critically on mathematical infelicities, except where these seem to lead to undesirable practical advice.

1. The authors' sampling plan, and other proposed plans for sampling inspection of meters, seek to give the operating company the information it needs about the condition of its meter population, and to do so with less inspection effort than the traditional system of 100% inspection after a fixed term of years. A properly designed sampling plan can estimate the mean error of a population of meters, and how closely individual meters are clustered about the population mean. Sampling inspection can do these things more efficiently and at lower cost than the traditional method. It can be objected that sampling inspection does not meet another purpose of regular inspection, that of protecting the individual consumer against the risk of many years of overcharges, based on a fast meter not discovered by regular inspection. This point could be discussed at great length, and no doubt it will be before sampling inspection of meters is adopted generally. It should be noted that, as shown by Fig. 3 of the present paper, older types of meters are rapidly going out of service. Once the class I and class II meters are gone, the case for 100% inspection will be greatly weakened.

2. The authors are certainly correct in saying that, when the distribution of meter errors in the population is known, the Variables method permits smaller sample sizes than the Attributes method for equal accuracy of estimation. The advantage of the Variables method is not, however, quite so great as Table II suggests. The ratio n_2/n_1 compares two sample sizes, a sample of size n_1 for estimating percentage defective and a sample of size n_2 for estimating mean error. The fact that n_2 is much smaller than n_1 is partly due to the fact that mean error is much easier than percentage defective to estimate from a sample. It follows that the values of n_2 given in Table V are not adequate for estimating percentage defective with the required precision. Values of n_2 needed for estimating percentage defective in a normal population can be found in reference 1.

3. A further weakness of sampling inspection by variables, not much discussed in text books, is the lack of robustness of the tests and confidence intervals for small proportions defective. Statisticians speak of a statistical procedure as being "robust" when its validity is only slightly affected by moderate departures of of the population from its assumed form. In the present case the critical assumption for the Variables method is the normal distribution of meter errors. If the proportion defective is small, moderate departures from a normal distribution may lead to serious errors in estimating the percentage of defective meters. Calculations carried out by Ontario Hydro for a population of Canadian meters show that, for this population at least, estimating proportion defective by the Variables method gives poor results.

4. In Table III of the paper, three different stratified sampling schemes are given for selecting a sample from each class of meters. A fourth scheme which has some advantages over those given in Table III would be an "optimum allocation" stratified sample, as developed by Neyman and described in reference 2. This method would give larger proportionate representation in the sample to the less accurate meters. The method can be generalized to allow for possible higher unit inspection costs of rural meters, and will show how to get specified accuracy at minimum cost.

References

1. Statistical Theory with Engineering Applications, A. Hald. John Wiley & Sons, Inc., New York, N. Y., 1952.

2. Sampling Techniques, W. G. Cochran. John Wiley & Sons, Inc., New York, N. Y., 1953.

A. J. Polaneczky (The Franklin Institute, Philadelphia, Pa.): The subject matter in the paper is of nationwide importance and undoubtedly will receive widespread attention. Unfortunately, we disagree with much of the philosophical presentation; we feel that the misapplication of statistics to the problem at hand invalidates a good deal of the paper. A discussion of the major points of disagreement follows:

1. The paper claims that sampling procedures sacrifice accuracy for economy. This is not necessarily so. There are instances when sampling gives more reliable results than 100% inspection. See page 9 of reference 7 of the paper.

2. The paper claims that the use of sampling presupposes a knowledge of the distribution of items in a universe. This knowledge would be helpful, but is not necessary.[1] The Binomial theory, as used in the paper, is nonrestrictive on knowledge of the universe distribution.

3. There are "Variables" methods, other than the one used in the paper, which do not demand that the universe be normally distributed. For a simple illustration see pages 55 and 56 of reference 7 of the paper, for a discussion of the Tchebycheff and Camp-Meidell inequalities.

4. The definition of a "Scientific Sample" is too restricted. There are occasions when a "biased" sample can give more reliable information than a random sample of the universe. See reference 2, with particular attention to Stratified and Cluster Sampling.

5. The equation developed for the relative size samples required by the Variables and Attributes plan is meaningless. The relative sample size is based on a comparison of confidence limits for the fraction defective calculated using the attributes method, with the confidence limits for the average registration (universe mean) based on the variables method. The relative sample size should be based on a comparison of the confidence limits for the fraction defective calculated using the attributes method with the confidence limits for the fraction defective based on the variables method. For the mechanics of calculating the fraction defective using variables data, see reference 3.

Unfortunately, the incorrect comparison, noted above, invalidates much of the latter sections of the paper.

6. It is not possible that the distribution of items in a universe, or sample thereof, be normalized by taking a sufficiently large sample. It is only the distribution of sample averages which is normalized.

7. The calculation of the confidence range, based on engineering experience, as illustrated by the authors (±2,890 meters out of 440,000) is unkind to the meterman. The number of defective meters was 5,780. With any kind of past history, the estimated confidence limits should be pretty good. The actual confidence limits, though, cannot be definitely tied down, unless more of the past history data is made available.

8. Again, in the Appendix of the paper, the universe and sample need not be normally distributed to make use of an attributes plan.

9. Equation 3 is incorrect. Sigma should be squared and the radical removed from n_2. The equations presented are concerned with confidence limits of the mean. As such, if the sample size is sufficiently large (>30), the restrictions on normality can be dropped.

References

1. Non-Parametric Statistics for the Behavioral Sciences, Sidney Segal. McGraw-Hill Book Company, Inc., New York, N. Y., 1956.

2. Sampling Techniques, William Cochran. John Wiley & Sons, Inc., New York, N. Y., 1953.

3. Statistical Theory With Engineering Applications, A. Hald. John Wiley & Sons, Inc., chap. 11, 1952.

C. L. Lucal (Ohio Power Company, Canton, Ohio): The authors are to be commended for presenting this approach to an evaluation of the accuracy of whr meters in service. As a result of the method, data obtained from tests on a small number of meters, properly selected, are the barometer for the performance of all meters of the same type in service.

Investigation of the accuracy of whr meters by classes was started before statistical sampling was presented to meter engineers. The data obtained were based on 100% samples of geographical areas. The number in the samples amounted to 2% of the meters in service.

In the work that has been done, all data were based on full load results. Tentatively, the limit of acceptability is that less than 2% of the sample may be outside the limits of 98% and 102% accuracy.

No attempt had been made to segregate urban and rural data, but it was evident that meters in areas of large rural content showed poorer performance than those in strictly urban areas. This was emphasized by the fact that the meters in classes IV and V were excellent when compared to those in classes I, II, and III. Rural and urban data were segregated in subsequent 100% samples of geographical areas. These data confirmed former evidence that meters in classes I, II, and III were affected to a much greater degree in rural areas than the same class of meters in urban areas, and is attributable to lightning exposure.

The performance of classes IV and V was not significantly different for the rural and urban areas. This is an indication that the surge-resistant magnets were not affected by lightning exposure in the rural areas.*

It should be pointed out here that all data in the foregoing discussion were obtained from meters that had been in service for 8 years. From these data it was concluded that:

1. Only meters having surge-resistant magnets should be installed in rural areas.

2. The test schedule for meters in classes I, II, and III could not be extended beyond the present 8-year schedule.

3. The performance of classes IV and V indicated that the test schedule could be extended.

4. The conclusions of items 2 and 3 were not compatible with area testing by the mobile laboratory method and would compel adoption of a due-date method, the cost of which might offset savings derived from a lengthened schedule for classes IV and V.

5. Economies in periodic schedules could be made by extending the schedules, but the most benefit could be attained when classes I, II, and III had been eliminated.

6. Periodic testing did not eliminate the cost of testing 96 out of 100 meters to find the 4 that needed attention.

Although the 100% sampling method afforded valuable data, it lacks the credence of the statistical sampling method which can be proved mathematically. This is important, for the statistical method has been found acceptable in many regions, whereas the data obtained through the 100% sampling method lacked the stature for acceptability. The statistical method also affords the probability that, in the long run, greater economies can be effected than by extension of periodic schedules.

When statistical methods are used for quality control of manufactured products it is assumed that the universe is homogeneous, i.e., that all the tin cans coming from one machine in a day's run are the same size, shape, color, and the last one made that day can be no older than the first by more than 24 hours, so time is not a factor. However, time is perhaps the most important factor when statistical sampling is used to draw inferences about the sustained accuracy of meters in service. This observation resulted from statistical sampling tests that were made.

It must be recognized that the average length of time in service since last date of test for meters in a statistical sample taken from a universe that has been on an 8-year schedule will be 4 years. It is for this reason that the performance data obtained by the statistical sampling method appears, on a percentage basis, to be better than that obtained by the 100% sample in a geographical area, the word "better" meaning that a smaller percentage were outside the 2% acceptability limit.

It seems necessary, therefore, to make an analysis by class of the meters that were outside the acceptability limit versus the length of time since last test. For classes I, II, and III about 1% were outside the limit at the end of 1 year, about 4% were outside the limit at the end of 8 years, and about 7% at the end of 11 years. This is evidence that there is a relation between accuracy and the years since last test.

A similar analysis for class IV shows that about 3% were outside the limit at the end of 8 years, a matter that should cause concern and indicates that, in the future, this analysis should be carried still further in that the data for each manufacturer's meters in the class should be analyzed.

The data for class V show excellent performance for an 8-year period. Here, again, the same precaution of detailed analysis is recommended in the future for this class.

Statistical sampling is a policing action regarding sustained accuracy over protracted lengths of service, and only by detailed analysis can the effects of time on a particular type and make be evaluated.

Once it has been established that a certain class of meter will not retain its accuracy for more than an 8-year period, it should not be necessary to include such meters in a statistical sample, thus saving the cost of testing such meters. It is presumed that these meters are placed on a regular schedule or are replaced by modern meters.

A saving can be effected by the statistical sample method through reduction of personnel and equipment, but the greatest benefits can accrue only when all the older classes of meters have been replaced with those which have inherent stability over a period of 30 to 40 years.

Referring to item 4 of the first heading under "Conclusions" in the paper, the statement should be qualified in that rural meters in classes I, II, and III show evidence of higher frequency of error than classes IV and V when compared to urban meters. Under item 5 of the same heading, the last clause of the last sentence would be more realistic if written, "1/7 of which will be tested in each of the next 7 years." Under this same item, it is difficult to agree that the 100% sample in a geographical area tells us nothing about the other 87½% of meters on the system. It is true that the 100% sample tells us the condition of the meters that have been in service for 8 years; we know, too, that the meters were within the accuracy limits when they were first installed. Consequently, it is a logical conclusion to feel that meters less than 8 years in service will have accuracies poorer than when they were originally installed, but in no instances poorer than at the end of 8 years. In other words, the deviation from original accuracy limits is a progressive situation. The only exception to this conclusion might be for a modern meter which it is believed will have a sustained accuracy for a period of 30 or 40 years.

W. C. Harrison (Virginia Electric and Power Company, Richmond, Va.): The authors have presented an excellent treatise on the use of statistical mathematics in the selection of whr meters for test.

For many years, the art of metering has been a highly specialized one, and many metermen have not had the opportunity of becoming acquainted with the statistical methods that have been employed in many other fields to solve test problems. To the meterman, upon first glance, the applications as stated in the report would indeed seem to be complex, and the initial reaction may be one of discouragement with the resulting failure to pursue the subject further. It is to be pointed out that even a brief insight into the use of statistical mathematics will prove a revelation for it will be found that all of the equations and calculations that might be employed will tend to prove that the judgments based on long years of experience with meters will be correct.

The use of such a plan as outlined for the selection of meters for test may not prove applicable where there are only a few meters to be considered, for the greatest benefit will be derived by those who are confronted with the enormous problem of caring for many thousands of meters. However, those with only a few meters could profit by a working knowledge of the sampling procedure for it will prove useful in many ways.

The domestic whr meter lends itself admirably to the sampling techniques. They are found in great numbers, their types are few, and their applications are similar to such an extent that the results obtained within one area of the country would even be useful for study purposes in other sections of the country. In fact, were this plan generally adopted, the interchange of information between groups would prove most valuable.

It is suggested that the meterman himself does not necessarily have to be an advanced mathematician, for one can usually find among his fellow engineers a mathematician who could quickly set forth the problems and the equations that may be required in the specific case. Indeed it is best that one, not too familiar with meters, be permitted to set up the basic formula for he would not be biased by knowledge accumulated from years of experience with meters. On the other hand, the meterman, knowing the characteristics of the various types of meters, would derive real pleasure in analyzing the mathematical proofs that his decisions through the years have been reasonably accurate.

G. A. Palmer (New England Power Service Company, Boston, Mass.): The authors have presented a very useful paper on the control of meter accuracies by statistical analyses of sample tests. The subject is a very live one in many electrical utility companies since the method holds great promise both for economy and for a gradual but significant improvement in meter performance. In the New England Electric System we have studied the sample test procedure for some time, with trial runs in two operating divisions. Generally our approach and our conclusions agree with those of the authors. Naturally there are differences, but no major ones, in details of procedure and in approach to the justification for this test method.

A difference in emphasis lies in our belief that sample testing can be justified by the fact that if proper action is taken on the results of the sample tests, meter accuracies can be improved over those achieved by periodic testing, regardless of how heavy the periodic test schedule may be. Of course the expected economies are welcome, but they are not the chief sanction for this procedure.

To comment on this paper is to comment on the subject of sample testing itself, as applied to whr meters. Accepting the procedure in its broad outlines, and accepting the mathematics which confirm the procedure, there remain some fundamental questions which should be raised. The answers to these questions should support the validity of the program rather than undermine it.

The aims of a modern test program must be that

1. Good meters are accepted with a minimum of effort.
2. Poor meters are corrected or replaced.

The question is: How do we find the poor meters? Let us not raise the question here as to the exact tenth of a per-cent error that is the dividing line between good and poor meters.

The authors answer, by a statistical approach. A sample test of the whole universe of meters analyzed by age, location, or, if necessary, by additional factors, will tell us which meters contributed more than their share of the total error.

There is no question but that a sample test can be used for quality control of any homogeneous group such as can be found on the production line of a factory, whether the product be meters or billiard balls. However, it does seem to me that certain modifications in thinking about sample testing must be made when the population is that of meters in service.

A company maintaining domestic whr meters on an 8-year schedule has performance data for 25% of these meters, collected in any 2-year period. Although in a number of places the authors admit that these performance data have given them information, they tend to disregard this information because the meters have not been selected according to the laws of random sampling. I do not believe that we are justified in throwing away the data we have so conscientiously collected over these many years. Mathematically it may be true that a 100% sample from one eighth of the total meter population does not justify any statistical inference about the remaining seven eighths. This implies that we are studying unknown devices under unknown conditions. Let us not adopt in our philosophy hypotheses that we do not believe in our everyday working life.

Every meterman knows that there have been improvements in meter design, and that the 1920 meter is not as good as the 1950 meter. Additional statistics are not required to establish this. We have 25 years of such statistics. In many types of meters we have already taken 100% sample. What more can we ask?

There is another aspect of the procedure of sample testing of whr meters which seems to me fundamental and which does not appear covered in this paper.

Professor Wilks[1] points out that in any sampling procedure correct conclusions depend not only on the method by which the sample is selected but on the process by which information is obtained on the sample once it is selected. Most of the emphasis so far has been placed on Stage I, the method of selection. But to quote Professor Wilks:

> "Stage II is the more basic of the two stages in the sense that it is at that stage when the degree of accuracy and precision (sometimes called validity and reproducibility) of the actual measurements themselves must be established. These cannot be established by merely assuming that the measurement obtained in a single test represents what we are usually tempted to think it represents. If measurements are repeated, variability is found and usually some of this variability is attributable to assignable causes (that is, causes which can be located) which may not have been suspected initially. Stage I sampling laws depend on making use of measurements whose accuracy and precision has been established at the stage II level."

This leads us to ask: When testing is done under differing conditions, as in a large company, with several test locations, how can it be determined how much variability in results is due to variability of meters, and how much due to variables in test equipment, technique, or personnel? A severe control of all factors is necessary, but it may be found that such control is difficult to maintain. Even with this control it must be admitted that variations in meters and test equipment make exact reproducibility of test results practically impossible. This seems to me particularly important when using the method of Variables.

In any statistical analysis the first step describes the universe and the second step proposes action. Using the method of Attributes, with a comparatively large sample, bases for corrective action, for acceptance or rejection, do not appear to me to be too difficult to establish. If our limits are ±2% the simple fact that a meter is or is not within these limits is sufficient.

With the method of Variables, however, the small sample calls for close analysis. I question whether the indicated degree of error is sufficient. I believe it is necessary to know the departure from previous calibration, that is, how much has each meter changed. Perhaps how much change per year.

It is impractical to adjust all meters to perfection, but with a sample as small as that stated to be sufficient for the method of Variables certainly the possibility that some meters were installed 1/2% slow and others 1/2% fast cannot be disregarded. Whether the meter $1^1/_2$% slow has changed 1% or 2% since last calibration should be of interest to the analyst. This requires reference to past history.

Referring again to Prof. Wilks, if some part of meter variability is attributable to assignable causes, should not this fact be considered in any analysis of the sample test results? If so, it is necessary to know in detail the history of the meter in service.

The question is, are we justified in using the method of Variables for determining performance of whr meters in service?

I should like to repeat that on the general program and on the conclusions reached we are in wholehearted agreement. I should like to thank Mr. Dwon and Mr. Morris for establishing the adequacy of a sample meter test program as a substitute for the inefficient and cumbrous periodic meter test schedule.

REFERENCE

1. **Wilks.** *Proceedings*, American Society of Testing Materials, Philadelphia, Pa., 1948.

L. M. Keever (North Carolina Utilities Commission, Raleigh, N. C.): In this paper the authors have dealt, in a very forthright manner, with a subject which should be of great interest to all those who have to do with single-phase whr meters. They are to be congratulated for making such a fine contribution and bringing this information in a very concise manner to many who can use it to good advantage.

Most of us have been aware for a number of years that a lot of time and energy is being wasted on testing, on a periodic basis of 5 or 8 years, these modern whr meters which are now being produced. One solution to this problem is a longer period between tests, but the question immediately arises as to what to do about the older types of meters which ostensibly do not have the qualities of sustained accuracy over long periods and with large overloads. Longer periods between tests obviously is not the logical answer.

We believe that sampling based on sound statistical principles and properly administered will give the utility and the regulatory body concerned a more intelligent evaluation of the condition of the utility's meters than can be gained by the present method of periodic testing. Analysis of the data from sampling tests will focus attention on the types of meters which have poorer performance characteristics and will help the meter department to convince management of the necessity of early retirement of these meters.

The authors have taken the position, and justifiably we think, of emphasizing the saving in meter test expense on the part of the utility which will ultimately be effected by the use of sampling in lieu of periodic testing. This is unquestionably an important feature of this system. We cannot overlook, however, the advantage of having rather accurate knowledge of the performance characteristics of all the various meter types which we have in service in a utility system. In what other way could we so accurately determine the susceptibility of the age group I, II, and III meters to the effects of exposure to lightning surges as is demonstrated in the authors' Table I. The much greater percentage of these age groups which are more than 2% fast in rural areas as contrasted to urban areas is emphatic when the small difference in these percentages is noted for age groups IV and V. It has been well known to all of us that the newer types of meters have characteristics of design which make them less susceptible to the effects of lightning surges, but it is only when we review the data on sampling tests that we have this fact brought so forcibly to our attention.

It has been assumed in general, as well as by the authors in this paper, that in the various age groups of meters which they have used for classification all manufacturers' types in each age group perform equally well. While this assumption is true in a general way, it has been found by analysis of sampling test data by meter types that specifically this assumption is subject to error for age groups I, II, and III. For example, if the average performance of a given age group indicates that 10% of this age group are more than 2% fast it is possible that this percentage is affected heavily by the poor performance of one type of meter in the group. On the other hand if another age group shows a large percentage of slow meters, this poor showing could be influenced strongly by a single type in this age group. Such discrepancies are not apparent to any appreciable degree in the test data of age groups IV and V.

We feel, therefore, that if a realistic approach is made in the analysis of these test data such analyses should be made by meter types rather than by age groups.

We further feel that a more accurate determination of the condition of all the meters in a utility system can be made if the sampling can be done on the basis of meter types rather than by age groups.

There is one point in connection with the substitution of scientific sampling of meters for test instead of the periodic system of testing on which the authors have not touched. This is the manner in which the system will be administered. Some standard of performance must be adopted whereby all meter types or age groups whose accuracy is shown by the sampling test data to be equal to or better than the standard will continue on sampling. Other types or age groups whose accuracies do not meet the standard will be tested on an accelerated periodic basis. In this group will be some types whose performance is sufficiently poor that they should be retired and replaced by modern types of meters. By this method if any group of meters after recalibration and modernization shows an improvement in operating accuracy to a point where it meets the standard established, then this group is placed on sampling rather than periodic testing. Similarly if any group in successive sampling tests falls below the standard such group would be placed on the accelerated basis of periodic testing.

It should be emphasized that the savings in meter testing costs envisioned by the authors will not be realized immediately upon the institution of this sampling procedure. During the first years of such procedure the accelerated periodic testing may require the testing of as many meters as are now being tested under normal procedure. Only after all the inherently inaccurate meters in the system have been eliminated will the full benefits of the sampling method be realized.

Larry Dwon and J. A. Morris: The value of a technical paper is appreciably enhanced by such honest and open discussions as have been advanced by F. H. Busch, W. C. Harrison, L. M. Keever, C. L. Lucal, G. A. Palmer, A. J. Polaneczky, and J. G. C. Templeton. Each of these gentlemen is recognized for his experience and knowledge of domestic meters and/or statistical sampling procedures.

The authors' presentation was prompted principally by the thought that perhaps there are some expert meter men who might not be adequately familiar with statistical methods and would, therefore, tend to shy away from investigating them. Thus, an example of how this powerful tool could be used seemed appropriate. This paper is a condensation of an internal report, which had for its basis these very concepts with relation to meter personnel in the AEP system. Experience with the report seemed to indicate that similar personnel in other areas of the country would welcome such a treatise. Therefore, it should be realized that this paper did not intend to cover all aspects of theoretical statistical sampling procedures, and it is for this reason that several references to standard texts have been included. As a consequence, it was anticipated that the more important omissions would be brought out by the discussers, and for this we thank each of them for having done so.

Particular attention is directed to Mr. Busch's remarks, which seem to cover most of the topics discussed, and in effect would constitute an ideal closure. It is well that Mr. Busch emphasized the advisability of learning fundamentals and principles of statistical sampling because they are rapidly becoming important tools in engineering. Mr. Busch's reference to the alternate approach of using probability statements instead of confidence levels and intervals is well taken, and certainly such approaches should be used by those who believe that they are more easily understood. Mr. Busch also makes very good reference to the many factors that must be evaluated in arriving at minimum accepted quality levels. Here is where statistical sampling may prove to be a valuable tool to help learn more about factors affecting meters in service, rather than simply to accumulate a mass of data. Statistical sampling methods may be used not only to determine meters for test, but also to excerpt from mass data information which may be considered too troublesome to obtain when faced with the review of all of it.

Mr. Templeton cautions us that the sampling inspection does not "protect the individual consumer against the risk of many years of overcharges based on a fast meter, which may not have been discovered by regular inspection." This is an obvious difficulty. The full value of statistical sampling procedures will not be derived until a population of meters is in existence which is mostly free of inaccurate meters. Mr. Templeton very accurately points out that the Variables method will not result in as small a relative sample as the n_2/n_1 ratio suggests in Table II. This point was discussed briefly in the paper and covered by the statement, "It should be noted, however, that the above formula exaggerates the n_2/n_1 ratio because the best attributes sample is not utilized as explained previously." Mr. Templeton points up a weakness of sampling inspection by variables, which should be taken into account. However, it would still appear that other advantages of the method outweigh the disadvantages. Mr. Templeton's reference to an "optimum allocation" stratified sample is a good one. As Mr. Busch pointed out in his discussion, there are many sampling plans that can be made, each having some particular advantage and disadvantage. It is up to the meter engineer and the statistician to determine what is best for the particular situation.

Mr. Polaneczky favored us with a discussion which we are happy to include because it points out quite positively that there is more to statistical analysis than what this paper had as an objective to show. The authors believe that Mr. Polaneczky did not appreciate that our paper did not intend to delve into the elegant phases of statistical philosophy. Our attempt was to acquaint an audience of meter men, who probably know little about elementary statistics, with this valuable and powerful tool in a specific field of application. In our attempt to meet the needs of this audience as simply as space limitations permitted, we may not have stated precisely some of the more philosophical facts about the subject. If we did not do an adequate job, the error certainly should be called to the reader's attention. However, we cannot agree that the points raised by Mr. Polaneczky are so badly distorted and misapplied, as he states. The authors wish to call the reader's attention to Mr. Polaneczky's text references, which deal with the deeper concepts that will certainly have to be taken into account as statistical procedures are applied to the meter testing problem.

Mr. Lucal, to whom the writers are indebted for much of the actual data used in the paper, points out that investigation of the accuracy of whr meters by classes was started before statistical sampling was presented to meter engineers. This was done on the basis of 100% samples in small geographical areas. This is acknowledged and merely confirms the value of sampling procedures. Whereas Mr. Lucal was able by such methods to obtain data about meters in a specific location, the same kind of information could have been obtained by statistical methods; but whereas in his situation he could conclude really only that the findings pertained to that geographical area, the statistical method would have allowed conclusions that pertained to the entire universe of meters in the company. Particular attention is directed to Mr. Lucal's very illuminating conclusions regarding meters. Mr. Lucal also makes mention of the homogeneity of the universe as used in quality control of manufactured products. Statistical methods can be used with nonhomogeneous universes, as have been pointed out by Mr. Templeton in his reference to statistical sampling.

Mr. Harrison cautions that the statistical sampling procedure may not prove applicable where there are only a few meters to be considered. This certainly is a worthy observation because statistical methods give the most help when they are applied to a mass of data from which information is desired at less expense, but at the risk of some error. Mr. Harrison brings out a very good point regarding the establishment of methods of interchanging information among groups throughout the country if standardized sampling techniques could be properly agreed upon. Mr. Harrison also makes a very good point when he suggests that the sampling plan might be worked out better by someone who is not familiar with meters in order to assure objectivity and an unbiased probe for correct information. However, the statistician would have to be guided by the meter expert in order to learn enough about the problem to properly establish the required sampling plan. This is essentially what the team of Morris and Dwon consisted of, the former a meter expert, the latter having some knowledge of statistical sampling procedures.

Mr. Palmer asked a very pertinent question, "How do we find the poor meters?" This, of course, can be achieved in many ways, but certainly cannot be done by the statistical sampling procedure. The latter will only determine the groups of meters predominantly in error and factors which may be instrumental in causing these errors. Once this is established, then the only way to determine all of the poor meters is to test all of them in that area or that group. As has been pointed out by other discussors and in the paper, the sampling procedure will be most effective when the universe contains meters of insignificant inaccuracy. Mr. Palmer's reference to the authors' disregard of information obtained from performance data requires a comment. In the paper, we intended to stress that if information is obtained about meters located in a concentrated area, any inferences that are made from such observation, theoretically or practically, cannot be carried over to meters in another location of the company. It isn't that the authors would disregard performance data, but it isn't possible to draw a conclusion from a test at location *A* of a miscellaneous group of meters which would be valid for another miscellaneous group of meters at location *B*. By no stretch of the imagination can one make such an inference. However, with an equivalent number of tests on meters picked at random from the entire universe, then similar information would be obtained, which then would be applicable to all of the meters on the system with certain error limits.

This is the main point that the authors were endeavoring to emphasize. Mr. Palmer brings out a very good point, however, regarding procedures of testing under different conditions in a large company. Certainly the Variables method depends on reproducibility of valid tests in all locations. Another good point that Mr. Palmer brings out is the impracticability to adjust all meters to perfection. The writers acknowledge that the subject of whether Variables or Attributes methods should be used is not a closed subject and, therefore, the question, "Are we justified in using the method of Variables for determining whr meters in service?" is not unreasonable. It will have to be determined by the needs through adequate discussions among the industries. We believe, however, that the Variables method, mainly because of its reduced sampling needs, should be the proper approach for our purpose.

Mr. Keever notes that is a realistic approach is made in the analysis of test data, it should be made by meter type rather than by age groups. The writers have no preconceived notions precluding other groups into which sampling information should be placed, and certainly meter type is one. The wide scope that is made available for analyzing the many factors affecting meters in service makes the area of meter testing exceptionally intriguing.

The writers are especially fortunate in having such prominent meter engineers and statisticians discuss this paper and feel obliged to thank all of the discussers for their excellent contributions and ideas.

Field-Strength Measurements in Parallel-Plate Precipitators

J. S. LAGARIAS
MEMBER AIEE

COMMERCIAL precipitators are faced with the problem of efficiently collecting dust particles of sizes ranging from hundredths of a micron to several hundred microns in diameter. This span of over ten thousand, in size, imposes a wide range of collecting conditions which make it difficult to accurately predict precipitator performance. The nature and magnitude of the forces acting on particles must be determined if one is to understand and improve precipitator design.

Electrostatic precipitators are generally tubular or of parallel-plate design. In the former, calculations can be made of the forces involved in solids removal. In parallel-plate precipitators, calculations are more difficult because the geometry does not readily lend itself to a solution for conditions wherein the space charge, dust loading, and field strength are variables. Because of the many conflicting factors present in commercial installations, it is necessary to isolate and study the many variables. This report describes studies made with experimental parallel-plate precipitator assemblies to determine the field strength distributions, the effect of dust loading, and to evaluate wire spacing in terms of known parameters.

Field strength was determined by measuring the charge on spherical steel spheres dropped through an electrical field. This technique has been used in several different forms. Pauthenier and Moreau-Hanot made many of the initial investigations of field strength in tubular electrodes.[1] In similar experiments, Penney investigated the effect of abnormal corona discharges on particle charge.[2] Cooperman used small metal spheres as missiles to measure fields and current densities.[3] In this investigation, emphasis was placed on using the dropping sphere technique to obtain field strength patterns in parallel-plate configurations and to determine the effect of dust loading on field strength.

Measurements were made with particles much larger than would usually be found in conventional precipitators but small enough so that their presence would not materially affect the field-strength dis-

---

Paper 59-206, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted November 15, 1957; made available for printing December 22, 1958.

J. S. Lagarias is with Koppers Company, Inc., Verona, Pa.

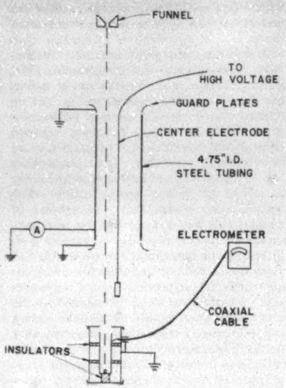


Fig. 1 (left). Charge measuring apparatus

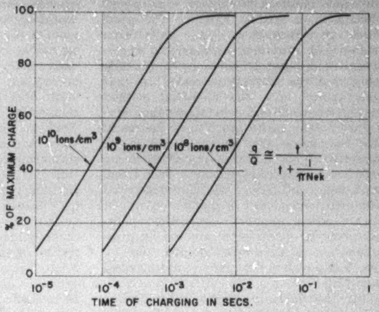


Fig. 2 (right). Ion density versus time on particle charge, field charging only

tribution in an 8-inch plate-to-plate geometry.

Description of Apparatus and Tests

Initial studies to determine charge as a function of field strength and particle size were made with a small precipitator section consisting of a 20-mil-diameter electrode centered in a 4.75-inch-inside-diameter steel tube open at both ends. The apparatus is shown schematically in Fig. 1. Spherical steel balls were dropped through the tubular-type electrostatic precipitator unit parallel to the discharge wires and the grounded cylindrical collecting electrode. Tests were made in still air at ambient temperature and humidity.

The spheres passed through the electric field between wire and collecting electrode and were caught in a Faraday cage. The charge on the ball was transferred to the cage and measured with an electrometer. The balls obtained their equilibrium charge since their fall was long compared to the charging time.

The per cent of saturation charge which a sphere will receive depends on the product of the ion density, N, and the charging time, T. The spheres were in the electric field region for 0.19 second. In the tests investigated, the ion density ranged from 1×10^8 to 1×10^9 ions/cm³ (centimeter) giving an NT factor and exposure time of 1.9×10^7 to 1.9×10^8. Fig. 2 illustrates the effect of ion density on per cent of charge on each sphere. The ion density and exposure time was sufficiently high so that the particles were fully charged for all practical purposes. The magnitude of the charge was a function of the field strength and the particle size.

All tests were made with the discharge electrode at negative polarity. At high field strengths, a noticeable deflection of the spheres occurred and it was necessary to move the cage accordingly. Spheres which were dropped too close to the collecting electrode and struck the walls before passing through the discharge electrode section, immediately gave up their charge and were effectively neutralized. All measurements were made of particles which passed through the precipitator section before striking the grounded electrode. Those particles which did strike the collecting surface were substantially neutralized.

It was assumed that the charge on each sphere represented the point of highest field strength through which the sphere passed. The deflection was always from a region of higher to lower field strength. In all instances, the sphere fell parallel to the discharge electrodes.

For balls greater than one micron, the charge on each sphere is proportional to the dielectric constant, the local field strength and the square of the ball radius.[1] This holds up to the size where the presence of a large charged sphere materially distorts the electric field. Field charging was the mechanism of charging, diffusion charging being negligible under the conditions involved.

Field-Strength Theory and Measurements

Field-strength measurements made in tubular precipitators can be used to check theory since calculations can readily be made. It is more difficult to calculate field strength for parallel-plate precipitators, however. In the present analysis, experiments were made with a tubular precipitator to check the reliability of the experimental technique. The field strength for parallel plate precipitators was then measured.

Equipotential surfaces in parallel-plate arrangements have been made using a probe technique but it is of interest to know the actual field strength since this is the driving force which is precipitating the dust from the gas stream.[4]

Tubular Electrodes

For concentric electrodes, with ionic space charge present, the distribution of the electric field strength, E, may be obtained from Poisson's equation written as

$$\frac{dE}{dr} + \frac{E}{r} - 4\pi\sigma = 0 \tag{1}$$

where σ is the volume space charge. The space charge is generated from corona emanating from the smaller of the two electrodes, and can be written as

$$4\pi\sigma = \frac{2i}{KrE} \tag{2}$$

where i is the current per-unit length of wire at a distance r from the wire since the ion current is

$$i = 2\pi r\sigma \times \text{ion velocity} \tag{3}$$

This solution to this equation as given by Thomsen is[5]

$$E = \sqrt{\frac{2i}{k} + \frac{C^2}{r^2}} \tag{4}$$

In the case where there is no current flow, this reduces to

$$E = \frac{C}{r} \tag{5}$$

which is the solution for the static field strength in concentric electrodes with

$$C = \frac{V}{\ln R_2/R_1} \tag{6}$$

and V being the potential difference be-

Vol.
Pt. 1
1959

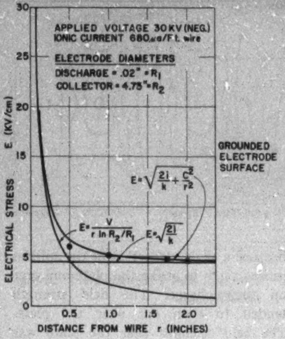
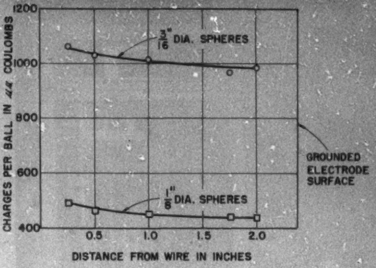


Fig. 3 (left). Field distribution in a tubular-type precipitator, cylindrical electrodes

Fig. 4 (right). Charge measurements in tubular-type precipitator, cylindrical electrodes, wire potential 30 kv, negative



tween electrodes. R_2 and R_1 are the outer and inner radii, respectively, of the electrodes.

For large values of r, the first term in equation 4 is large compared to the second and a reduction to the generally used simplified form is possible:

$$E=\sqrt{\frac{2i}{k}} \tag{7}$$

The field strength is essentially independent of the radius throughout most of the outer region of the annular space. In the course of the investigation, it was necessary to calibrate the equipment and Fig. 3 indicates the field strength as determined in this manner and as checked experimentally. Results of this type indicate the validity of measuring field strength by the dropping sphere technique.

Equation 4 is applicable close to the wire but reduces rather rapidly to equation 7 at a short distance from the wire. It is seen that an essentially flat field strength region exists between electrodes. Fig. 4 illustrates the uniformity of the field between electrodes as measured with 1/8-inch- and 3/16-inch-diameter spheres.

Effect of Dust Burden

The presence of dust or solid particles in the gas stream will distort the field-strength distribution still further because of the lower mobility of the dust particles. As developed by Rose and Wood,[6] Poisson's equation is

$$\frac{dE}{dr}+\frac{E}{r}-\frac{2i}{krE}-pES=0 \tag{8}$$

where p is the particle charging factor involving the dielectric constant, D,

$$p=\left[1+2\frac{(D-1)}{(D+2)}\right]$$

and S is the surface area of the dust particles per-unit gas volume.

The solution to this equation is

$$E=\frac{1}{r}\left[\left(\frac{i}{kp^2S^2}+C^2\right)e^{2pSr}-\frac{2i}{pSk}\left(r+\frac{1}{2(pS)^2}\right)\right]^{1/2} \tag{9}$$

Expanding and assuming that pS is small:

$$E=\left[\frac{2i}{k}\left(1+\frac{2pSr}{3}\right)+\frac{C^2}{r^2}\right]^{1/2} \tag{10}$$

which reduces to equation 3 when there is no dust present. For the case where there is dust present in a tubular electrode, the following equation may be used in the region near the plate:

$$E=\sqrt{\frac{2i}{k}\left(1+\frac{2pSr}{3}\right)} \tag{11}$$

It is seen that the presence of dust will increase the field strength assuming the discharge current is constant. The current, however, is not independent of the dust loading.

If spherical particles of mean radius, r_0, are assumed, then equation 11 for tubular electrodes can be written

$$E=\sqrt{\frac{2i}{k}\left(1+\frac{pLr}{\rho r_0}\right)} \tag{12}$$

where

L = dust loading per-unit volume of gas
ρ = density of the dust

The effect of loading on precipitator performance can be estimated in this manner.

PLATE-TYPE ELECTRODES

An examination of the field-strength distribution in parallel-plate electrode arrangements is of special interest since most commercial heavy dust precipitators are more difficult to calculate because of the complications of geometry.

Troost has shown[7] that if equation 7 holds, then the field strength may be given as

$$E=\sqrt{\frac{8iD}{kB}} \tag{13}$$

where

D = distance between wire and plate
B = distance between adjacent corona wires
i = current per-unit length of wire

for a parallel-plate arrangement. This assumption is based on a triangular field distribution and assumes no field exists outside of this region which is, of course, not the case. It is seen that for the condition where the distance between wires is four times the distance between wire and plate the field strength will be essentially the same as that for concentric electrodes.

The field strength of a plate-type arrangement was measured experimentally using the measurement of charge technique as shown schematically in Fig. 5. Steel spheres were dropped parallel to the discharge electrodes at various distances from the plates and discharge electrodes and the field strength determined by the magnitude of the charge on each sphere. Each measurement was the average of ten or more measurements giving good reproducibility. Variation in charge between individual measurements was approxi-

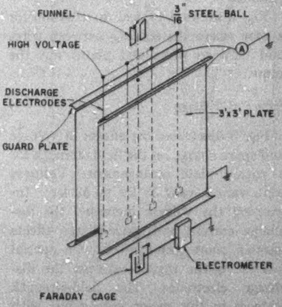


Fig. 5. Parallel-plate precipitator

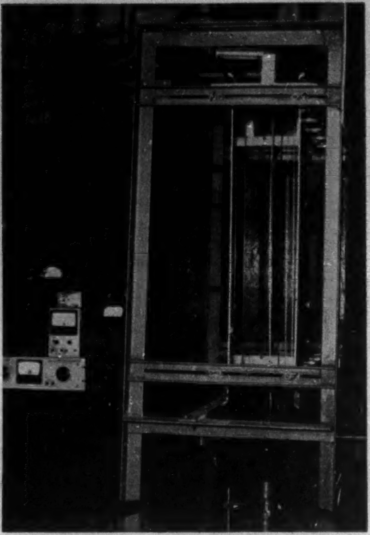


Fig. 6 (left). Parallel-plate field measuring equipment

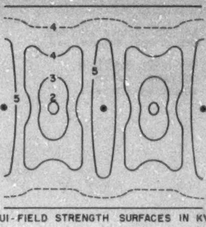


Fig. 7 (right). Field-strength distribution in air. Wire-to-wire spacing, 4 inches; plate-to-plate spacing, 8 inches; voltage, 53 kv, negative

mately ten % but the average values were reproducible within 5%. Day by day reproducibility was consistently good after edge effects were minimized. The equifield strength surfaces were obtained by connecting points of equal field strength. The test apparatus used is seen in Fig. 6.

The charge on the spheres was taken to represent the maximum field strength encountered which would be at the point where the sphere entered the influence of the electric field. That this is representative of the field is shown by the excellent agreement in the tubular precipitator arrangement between the field calculated using equation 4 and the experimental results.

Fig. 7 shows a typical distribution for a case where the spacing between wires and between wires and plates was the same.

Effect of Voltage

Fig. 8 illustrates the effects of voltage and space charge on the field strength for a parallel-plate arrangement. Voltages were varied from 33 kv to 53 kv. Increasing the voltage increased the discharge current and space-charge effects became more pronounced. Of special interest is the region between the discharge electrodes. The field-strength surfaces in this region spread out with higher current densities while the field region between wire and plate remains narrow. Field-strength lines were determined by connecting points of equal field strength. The dashed line srepresent field-strength lines near the plate. Data in this region were difficult to obtain because a slight deflection of the spheres caused them to strike the plate and give up their charge. The field strength tended to even out near the plate. The field strength near the plate was lower at points midway from the wires than directly opposite the wires.

Effect of Spacing

The effect of wire-to-wire spacing on the field strength is illustrated on Fig. 9. As the spacing between wires increases, there is a general spreading of the field from the mid-point between wires rather than from the wires themselves. A doughnut effect beings to show up as the spacing is increased.

Field Strength in the Direction of Gas Flow

To an entering dust particle traveling along a path parallel to the plates, the field strength will be a function of position varying with time and with distance

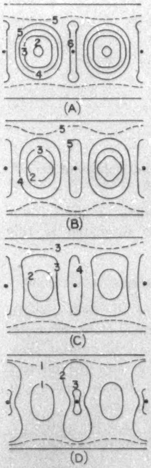


Fig. 8 (left). Voltage effect on field strength (kv/cm). Plate-to-plate spacing, 8 inches; wire-to-wire spacing, 6 inches

A—Wire potential 63 kv, negative
1 = 325 μa/ft wire
B—Wire potential 53 kv, negative
1 = 135 μa/ft wire
C—Wire potential 43 kv, negative
1 = 20 μa/ft wire
D—Wire potential 33 kv, negative
1 = 8 μa/ft wire

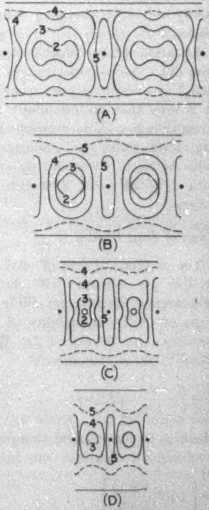


Fig. 9 (right). Effect of wire-to-wire spacing on field strength. Wire potential 53 kv, negative; plate-to-plate spacing, 8 inches

A—8-inch spacing
B—6-inch spacing
C—4-inch spacing
D—3-inch spacing

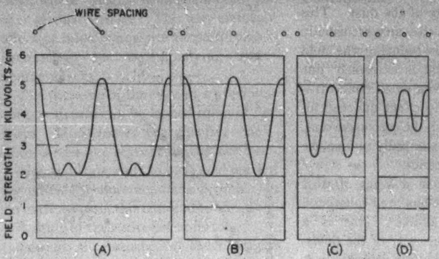


Fig. 10. Field strength 1 inch from wire. See subcaption Fig. 9 for identification

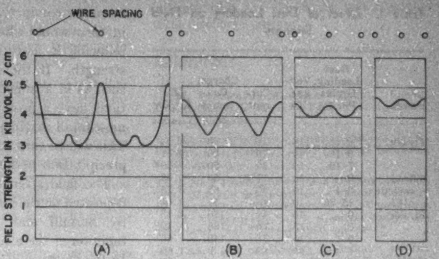


Fig. 11. Field strength 2 inches from wire. See subcaption Fig. 9 for identification

from electrodes. It is evident from Fig. 9 that the field strength directly between wire and plate does not change for the wire-to-wire spacings investigated. The field between wire, however, is highly sensitive to the spacing. Fig. 10 compares the field strength between wires at a distance of 1 inch from the wires. The peak field strength directly opposite the wires is essentially the same for all wire-to-wire spacings. The dips in field strength between wires, however, are much greater for the larger wire-to-wire spacings. The average field strength close to the wires is higher for the small wire-to-wire spacings. For high field-strength applications then, it is desirable to have the wires close together. This would reduce the reductions in field strength between electrodes.

Fig. 11 shows the field-strength variation with distance parallel to the plates at a distance of 2 inches from the discharge electrode. The dips between electrodes are not as pronounced as the collecting electrode is approached. This smoothing of the field still shows an increased average field strength for the wires which are closely spaced.

Wire Spacing on Current-Voltage Characteristics

As the spacing between wires is decreased, however, there is an interaction between electrodes in which the field emanating from a wire tends to suppress the current from adjacent electrodes. Fig. 12 shows the current-voltage characteristics of a flat-plate precipitator section as a function of wire-to-wire spacing. This would result in a decreased operating range in commercial precipitators and will make control of a precipitator with closely spaced discharge electrodes a little more difficult than with more widely spaced wires.

Fig. 13 illustrates the manner in which the total current/sq ft (per square foot) of plate surface area is affected. In spite of the wide range of wire-to-wire spacings tested, the current/ft of surface area is essentially independent of the wire spacing.

At still lower spacing, however, on a given precipitator length, the interaction between wires will reduce the current at a greater rate than the increase resulting from using more wires. The total current per-unit plate area will thus be reduced.

As the spacing is increased above 8 inches, very little additional benefit is achieved. The increase in current per wire will be slight compared to the increased area which each wire must service again reducing the current per-unit plate area.

In practice, the minimum spacing to be used is dictated by a number of factors: the cost of installing and maintaining additional electrodes, the increased steepness of the discharge characteristics leading to a smaller operating range, and the advantages to be gained therein. In general, however, increased field strength may be obtained by reducing the spacing between electrodes as indicated by Fig. 10 and 11. Since this is the driving force tending to remove particles from the gas, a high electric field is usually desirable. Each precipitator installation has its own unique problems, however, which means other factors such as type of dust, loading, resistivity, etc., must also be considered.

Effect of Dust Burden

The field-strength surfaces shown in Figs. 7–9 were taken in a dust-free atsphere. The question arises as to the

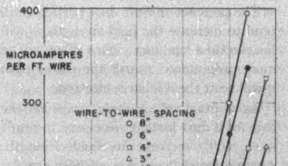


Fig. 12. Effect of wire-to-wire spacing on discharge current. Plate-to-plate spacing, 8 inches

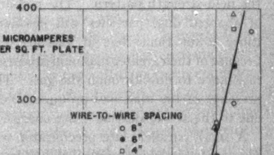


Fig. 13. Effect of wire spacing on current to plates. Plate-to-plate spacing, 8 inches

Table I. Effect of Dust Loading on Field Strength

Powder	Dust Loading, Grains Per Ft^3	Voltage, Kv	Current	Charge, Coulombs	E, Kv/Cm
Flyash	0	10	10	45	0.54
	0–1	10	15	60	0.71
	7–10	10	20	50	0.60
Silicon carbide (3F)	0	10	17.5	53	0.63
	2–4	10	20	63	0.75
	15–20	10	13	54	0.64
Flyash	0	15	64	120	1.43
	0	15	76	125	1.49
	0–1	15	71	140	1.66
	1–4	15	88	140	1.66
	15–20	16	76	126	1.50
	10–15	15	90	130	1.55
Silicon carbide (3F)	0	15	72	130	1.55
	2–4	15	74	139	1.65
	10–15	15	68	126	1.50
Lampblack	0	15	76	135	1.60
	1–3	15	62	121	1.44
		15	Arced		
Aluminum oxide (1,000)	0	15	62	138	1.64
	2–4	15	64	125	1.49
Flyash	0	20	215	205	2.44
	0	20	185	216	2.56
	2–4	20	195	190	2.26
	8–10	20	200	205	2.44
Silicon carbide (3F)	0	20	190	204	2.42
	2–4	20	200	207	2.46
	8–10	20	200	209	2.48
Lampblack	0	20	175	203	2.41
	1–3	20	105	149	1.77
	1–3	20	130	171	2.03
		20	Arced		

Table IA

Dust	Density, Grams/Cm^3	Average Particle Size, Microns
Flyash	2.4	5.0
Silicon carbide	3.1	8.7
Al_2O_3 (aluminum oxide)	3.9	4.3
Lampblack	1.3	1.25

effect the presence of dust will have on the field-strength pattern. The mobility of charged dust particles will be many times lower than that of gaseous ions because of their greater mass and increased resistance to flow through the gas. The presence of the dust will produce a field due to the slower moving space charge.

A small parallel-plate precipitator was constructed in which controlled dust loadings could be obtained. The field strength at several points was determined using the dropping sphere technique. Five widely different materials were used as the dust burden. The results are tabulated in Table I. The most significant effect of dust loading is the manner in which the current is changed.

The primary effect of dust loadings was to increase the field strength because of the space charge generated by the slow moving charged particles. Other factors, however, were also influenced by the presence of the dust. The effective resistance of the air gap between electrodes for example, may be reduced depending upon the conductivity of the dust. This in turn would alter the current, usually reducing it and thus reducing the field strength. If the dust is extremely fine such as in a fume, even small concentrations have a profound effect and can almost totally extinguish the corona current. This will in turn substantially reduce the precipitation performance.

The field strength in a layer of dust deposited on the plate has been calculated by Sproull and Nakada.[8] This is a function of the conductivity of the dusts. In the presence of a high-resistivity dust, the field strength may be altered considerably. This investigation, however, concerned itself only with the discharge electrodes and grounded parallel plates, however, without the added complication of current flowing through precipitated dust. It was recognized that the field would be altered by the manne in which the current flows through a dus layer.

In a tubular precipitator, the field strength will vary along the wire depending upon the degree of loading. At the inlet, the field strength will be increased due to the dust if the current/ft of discharge wire is uniform. However, the current-limiting effect of the dust may reduce the field strength more than the increase due to the charged dust particles.

For a single point, equation 12 may be modified to apply to flat-plate units to

$$E = C_0\sqrt{\frac{2i}{k}+\frac{pLr_0}{r}} \qquad (14)$$

where C_0 is a constant depending on the geometry and the point of investigation.

The presence of finer dust particles will tend to increase the field strength as will heavier dust loadings. This effect will be more pronounced nearer the plates than it will be at the discharge electrode.

As a practical example, consider the case for a dust loading necessary to result in a 10% increase in field strength. Assume flyash having a specific gravity of 2.4, a mean particle diameter of 5 microns, and an 8-inch plate-to-plate spacing, then determine the field strength at a point 1 inch from the plate. For $p=2.8$, it would require a dust loading of 8 grains/ft^3 to result in a 10% increase in field strength. This assumes the current would remain constant under such conditions.

In actual practice, however, the presence of a dust load usually decreases the discharge current to the extent that a modification of equation 7 may be applied through most of the effective cleaning region. Fig. 14 (base on data from Table I) illustrates the actual measurement of field strength at a single point between varying conditions of dust loading. The solid line is drawn for the equation

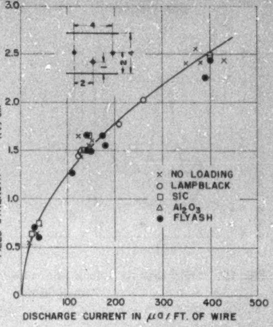


Fig. 14. Dust-loading effect on field strength in a parallel-plate precipitator, $E=0.70\sqrt{\frac{2i}{k}}$

$$E = C_0\sqrt{\frac{2i}{k}} \qquad (15)$$

where $C_0=0.70$ for the test point examined. It is evident that for all practical purposes the field strength depends more on the current than it does on the dust burden. In addition to making tests with varying degrees of dust loading, tests were made with widely differing particle size distributions and conductivity. Fig. 14 shows the field strength to be uniquely determined by the discharge current. The field strength is thus affected by the dust loading depending upon the manner in which loading affects the current. The additional effect of the charge carried by the dust was not found even at loadings as high as 20 grains/ft.^{3} It is likely, however, that for extremely fine fumes such as zinc oxide for example, the space charge carried by the dust may alter the field strength to a greater extent than indicated by current alone. For most commercial dust loadings, however, the presence of dust in the gas will vary the field strength only as it affects the discharge current.

Effect of Particle Size

The effect of particle size depends upon the area-to-weight ratio of the various size fractions. The surface area of each dust fraction is proportional to the weight of the dust fraction and inversely proportional to the square of its radius. Thus, the surface area present on 1.0 micron particles is equivalent to the area of 100 times as much dust by weight of 10 micron particles. The total effect of dust whose weight is negligible in the over-all dust burden. This may seriously affect or impair precipitator performance because of the accompanying space-charge effects.

The presence of the dust in a precipitator gas stream can be looked upon in two ways. It decreases the effective resistance of the gas separating the two electrodes and thus leads to increased current by making a more conducting path. In an extreme case, this can lead to the short-circuiting of the entire precipitator.

Second, the presence of the relatively slow moving space charge created by ions attaching themselves to the particulate matter, results in a field which is acting so as to impede the further flow of current. For a very fine nonconducting power it is possible to reduce the total current to a mere fraction of the original total current.

Formulas in this article are given in electrostatic units. Practical units are used in the curves and figures, however, for convenience.

Conclusions

1. Field-strength patterns may be determined using a dropping sphere technique in air and in a dust-laden atmosphere.

2. Field-strength patterns in a parallel-plate precipitator section indicate that increasing the discharge current increases the field strength between electrodes whereas the field directly between wire and plates is substantially unchanged.

3. An unusually high field-strength region has been found to exist in the region between discharge electrodes and is attributed to space charge.

4. The effect of wire-to-wire spacing has substantially little effect on the current-flowing per-unit plate area but the current voltage characteristics become steeper for the smaller wire-to-wire spacings.

5. A higher over-all field strength is possible with reduced wire spacing because the dips in field strength between wires are reduced.

6. The presence of dust loading in a precipitator will change the field strength as it affects the discharge current. When theory indicates surface area of the dust is a significant factor in the field-strength distribution in a precipitator, experimental data indicate that the significant factor is the total discharge current.

References

1. **M. Pauthenier, M. Moreau-Hanot.** *Journal de Physique et le Radium*, Paris, France, vol. 6, series 7, 1935, p. 257.

2. MEASUREMENTS OF ABNORMAL CORONA, **G. W. Penney, J. G. Hewitt.** *AIEE CP 57-87* (available on request).

3. A NEW TECHNIQUE FOR THE MEASUREMENT OF CORONA FIELD STRENGTH AND CURRENT DENSITY IN ELECTRICAL PRECIPITATION, **P. Cooperman.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, Mar. 1956, pp. 64–67.

4. THE ROLE OF CORONA DISCHARGE IN THE ELECTRICAL PRECIPITATION PROCESS, **H. J. White.** *Electrical Engineering*, vol. 71, Jan. 1952, pp. 67–73.

5. CONDUCTION OF ELECTRICITY THROUGH GASES, VOL. II (book), **J. J. Thomsen, G. P. Thomsen.** Cambridge University Press, Cambridge, England, 1933.

6. AN INTRODUCTION TO ELECTROSTATIC PRECIPITATION IN THEORY AND PRACTICE (book), **H. E. Rose, A. J. Wood.** Constable & Company, Ltd., London, England, 1956.

7. **N. Troost.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 101, pt. II, 1954, p. 369.

8. **W. T. Sproull, Y. Nakada.** *Industrial and Engineering Chemistry*, Washington, D. C., vol. 43, 1951, p. 1359.

Automatic Regulators with Self-Balancing Magnetic Amplifiers

W. A. GEYGER
MEMBER AIEE

Synopsis: This paper is concerned with the application of the self-balancing magnetic-amplifier technique in the realm of automatic regulators. It describes two typical examples of such arrangements: 1. A constant-voltage regulator with smoothed low-impedance d-c output to be used for charging of special types of storage batteries, and 2. a constant-light-intensity regulator with a barrier-layer-type photoelectric cell for instrumentation. Both regulators employ self-balancing magnetic amplifiers with infinite internal gain and have excellent performance characteristics with regard to the respective control problems. The paper includes some considerations on the associated constant-voltage and constant-current reference circuits.

IN SELF-BALANCING magnetic amplifiers[1–3] with infinite internal gain, "critical regeneration," the output-input characteristic is defined by the resistance values of the negative-feedback channel and is not substantially influenced by changes in magnitude and frequency of the power-supply voltage, large variations in load resistance, etc. If the input voltage or input current is constant, then the output voltage or output current is constant. In this case, the self-balancing magnetic amplifier acts as a high-performance automatic regulator, the output-input characteristic (voltage ratio, current ratio, transconductance, or transresistance, respectively) of which is solely determined by actual values of resistance components. As a result of the large amount of negative electric feedback, resistance coupled into the control-circuit loop, this type of regulator has excellent performance characteristics with regard to various control problems.

This paper describes two typical examples of automatic regulators with self-balancing magnetic amplifiers: 1. A constant-voltage regulator with smoothed low-impedance d-c output to be used for charging of special types of storage batteries, and 2. a constant-light-intensity regulator with a barrier-layer-type photoelectric cell for instrumentation. Because actual performance of such arrangements depends upon that of the associated constant-voltage or constant-current reference circuit, respectively, the paper includes some considerations on such circuits.

Constant-Voltage Regulator With Smoothed Low-Impedance D-C Output (Battery Charger)

GENERAL

Some early applications of saturable-reactor devices were concerned with charging of storage batteries under automatically controlled operating conditions.[4] Rated values of output voltage and output current may be varied within wide limits, in accordance with actual requirements. When smoothed d-c output is desired, it is generally possible to apply *L-C* filter circuits in the usual way.

Development of standard-voltage or

Paper 59-174, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 27, 1958; made available for printing December 9, 1958.

W. A. GEYGER is with the United States Naval Ordnance Laboratory, Silver Spring, Md.

The author would like to express thanks to Edward T. Hooper for his helpful co-operation in conducting the oscillographic investigations. Indebtedness is also acknowledged to Walter P. Christoph and Herbert B. Reed for their help and valuable suggestions during the course of the research leading to this paper.

standard-current circuitry for automatic regulators has taken two courses. The first involves the application of a voltage discriminator, i.e., a voltage-sensitive auxiliary circuit, for example, in the form of a Wheatstone-bridge network with two linear resistors and two nonlinear resistors (e.g., metal-filament lamps). This bridge network is balanced if the regulated voltage has its rated value. Otherwise, this bridge network is thrown out of balance and produces a polarity reversible output current, which corresponds in its magnitude and direction to actual deviation of the regulated voltage from its rated value. Of course, other types of nonlinear impedance networks may be used equally well for this purpose.

The second approach consists in comparing actual value of the regulated voltage with a standard voltage which is derived from a constant-voltage reference device. There are various possibilities for comparing these two voltages in a voltage-type or current-type differential circuit. In this case, too, the resultant polarity-reversible output current corresponds with regard to its magnitude and direction to actual deviation of the regulated voltage from its rated value. This approach based upon application of a constant-voltage reference circuit has been used in the arrangement described in the following sections.

The Specific Problem

It is common practice to operate magnetic-amplifier-controlled metal-rectifier equipment for charging of lead-acid storage batteries in such a way that the voltage is maintained at 2.15–2.18 volts per cell, and so that the rectifier normally supplies the load current and the battery draws only a small current sufficient to maintain the charged condition.[5] Because in this case, the battery supplies current only if the load requirements are in excess of the rated rectifier output or if a failure of the a-c power supply occurs, the life of the battery is exceptionally long.[6]

An actual problem consisted of providing a magnetic-amplifier voltage regulator with smoothed low-impedance d-c output to charge a special type of storage battery, either single-celled or with several cells in parallel connection. The requirements were a rated value of output voltage of 1.12 volts and a regulation accuracy of ±2% for load-current variations from no-load to full-load conditions (from 0–4 amperes), for ±10% voltage variations of the frequency-regulated 60-cycle power supply.

In conventional forms of automatic voltage regulators with magnetic-amplifier circuits, *L-C* filters are successfully employed for reducing the ripple voltage, i.e., the a c component of the d-c output voltage, to sufficiently small values. However, in the present problem, application of such filter circuits for reducing the a-c component of the highly nonsinusoidal magnetic-amplifier output voltage is practically impossible for very low values of load-circuit resistance (about 0.3 ohm with full-load-current conditions).

The main difficulty of the problem is that the waveshape of the output voltage of a high-gain magnetic amplifier using rectangular hysteresis loop core material is generally a truncated sine wave which consists of a fundamental component and an array of harmonics having relatively large magnitudes. Thus, very effective filter circuits are necessary to reduce the large a-c components of the magnetic-amplifier output voltage to sufficiently small values.

The specific problem, as outlined in the foregoing, has been solved by introducing an ordinary saturable reactor operating with rectangular-hysteresis-loop core material under forced-magnetization conditions[7,8] acting as a low-gain square-wave circuit (waveform converter) between a high-gain self-balancing magnetic amplifier and the load circuit. In this way, the essential difficulty of filtering highly distorted unidirectional output voltages has been removed, because application of a full-wave-rectified square-wave output voltage minimizes the requirements on filter components. Indeed, the artifice of employing an additional saturable-reactor circuit acting as a waveform converter has proved to be of decisive importance with regard to the solution of the problem. This approach, of course, may also be attractive in other cases, where it is desirable to reduce the a-c components of the output voltage and to minimize requirements on filter components.

Application of a self-balancing magnetic amplifier with infinite internal gain offers various advantages concerning the performance characteristics of automatic voltage regulators. However, because these amplifiers operate in principle under natural-magnetization conditions[7,8] the output-voltage waveshape corresponding to that of conventional forms of high-gain self-saturating circuits is highly distorted. For this reason, the artifice of introducing an additional saturable reactor circuit acting as a waveform converter is of particular importance. In this way it is possible to combine the specific advantages of both types of magnetic-amplifier circuits (high gain of self-balancing magnetic amplifiers, and square wave output of ordinary saturable reactors operating under forced-magnetization conditions) to obtain optimum performance.

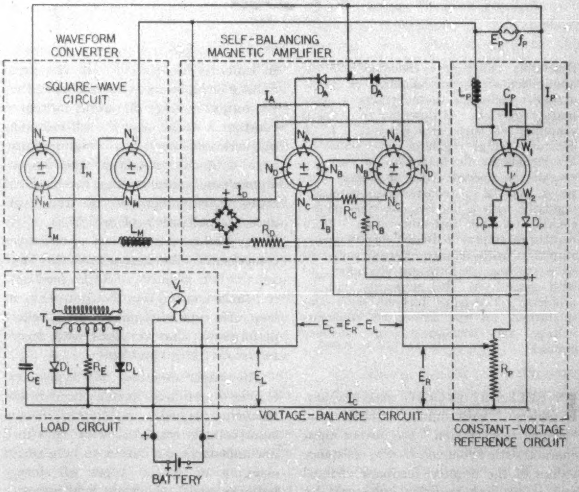


Fig. 1. Circuit diagram of the constant-voltage regulator (battery charger) with a self-balancing magnetic amplifier operating under voltage-balance conditions

THE REGULATOR CIRCUIT[7]

Fig. 1 shows a circuit diagram of the automatic voltage regulator, which has four main parts: 1. The constant-voltage reference circuit providing the constant reference voltage E_R. 2. The load circuit applying to the storage battery (or some other d-c load) the load voltage E_L to be regulated to have a constant value. 3. A self-balancing magnetic amplifier with the voltage-balance circuit (resultant control voltage $E_C=E_R-E_L$) and a full-wave rectifier in the output circuit. 4. An ordinary saturable reactor with smoothed control current I_M (choke coil L_M) acting as a waveform converter and producing the square-wave output current I_N).

The magnetic voltage stabilizer of the constant-voltage reference circuit is a parallel-type ferroresonant network consisting of a nonlinear reactor T_P with rectangular magnetic characteristic (Orthonol 2-mil tape core) and two separate center-tapped toroidal windings W_1, W_2, a capacitor C_P, and a linear reactor L_P with Permalloy-dust core. Potentiometer resistor R_P is supplied from a full-wave rectifier circuit with two silicon diodes D_P and permits adjustment of the reference voltage E_R to the desired value, e.g., 1.12 volts.

The self-balancing magnetic amplifier comparing actual magnitudes of the control-voltage components E_R and E_L has the following parts: 1. The self-saturating circuit with Orthonol 2-mil tape cores, gate windings N_A and series-connected silicon diodes D_A. 2. The bias circuit with d-c bias windings N_B and series resistor R_B provided for adjusting the operating point of the self-saturating circuit, i.e., actual magnitude of the quiescent current (output current I_A with no-signal conditions). 3. the d-c control windings N_C (resistance R_C) carrying the control-winding current I_C, which is substantially reduced to zero. 4. The d-c positive-feedback windings N_D with regeneration-control resistor R_D for adjustment of critical regeneration (infinite internal gain of the amplifier).

The waveform converter consisting of an ordinary saturable reactor with Orthonol 2-mil tape cores, series-opposing connected a-c load windings N_L, series-aiding-connected d-c control windings N_C and choke coil L_M is characterized by the fact that its d-c control circuit represents a very high impedance with regard to even harmonic currents. Thus, the saturable-reactor elements operate under forced-magnetization conditions[7,8] i.e., the alternating output current I_N has a nearly reactangular waveshape.

The load circuit converting the alternating square-wave current I_N into a smoothed unidirectional output current contains the following components: 1. The output transformer T_L with center-tapped secondary winding. It acts as a step down transformer (voltage ratio 60/(3+3) volts) and provides the necessary insulation between the a-c power supply (voltage $E_P=120$ volts ±10%, frequency $f_P=60$ cps) and the control-circuit components. 2. Two silicon diodes D_L providing full-wave rectification. 3. An internal load resistor $R_E=1$ ohm, which limits the output-current range of the regulator to a minimum value of about 1 ampere. Thus, even if the external load current drawn by the battery or some other d-c load is reduced to zero, the saturable-reactor square-wave circuit is still loaded by the constant internal load current flowing through the 1-ohm resistor R_E. 4. A 4,000-microfarad smoothing capacitor C_E, which substantially removes the remaining sharp-pulse a-c components of the full-wave-rectified load voltage E_L. 5. A moving-coil voltmeter V_L for indicating or recording of actual magnitude of the d-c load voltage E_L.

PRINCIPLE OF OPERATION AND PERFORMANCE CHARACTERISTICS

When drawing no external load current from the output terminals of the regulator, the magnitude of the output current is solely determined by the 1-ohm resistor R_E and the parallel-connected high-impedance voltmeter V_L. In accordance with the basic principle of voltage-balance-type self-balancing magnetic amplifiers with infinite internal gain, the reference voltage E_R (representing the signal voltage) is balanced by the load voltage E_L (representing the compensating voltage), and, with critical regeneration of the amplifier, the current flowing in the control-circuit loop is substantially reduced to zero. If this case, the internal output current of the 1-ohm resistor with voltmeter V_L is controlled by an infinitesimal d-c flux increment in the cores of the self-saturating circuit. It is to be noted that the balance conditions are not substantially influenced by changes of the characteristics of the amplifier components, fluctuations of the 60-cycle power-supply voltage E_P, et cet.

If now an external load resistor, e.g., a discharged storage battery, is suddenly connected with the output terminals of the regulator, an external load current is drawn, which increases the voltage drops across the secondary winding of transformer T_L and the rectifier elements D_L. As a result, during the transient operation of the regulator, the load voltage E_L drops in a certain amount, i.e., $E_L<E_R$, and a resultant control voltage $E_C=E_R-E_L$ of positive direction causes alternating output current I_A, smoothed unidirectional control current I_M, square-wave current I_N, and the d-c output voltage E_L to increase in such an amount that balance conditions are restored. On the other hand, if the external load current is suddenly reduced, then $E_L>E_R$, and a resultant control voltage $E_C=E_R-E_L$ of negative direction will cause currents I_A, I_M, I_N, and voltage E_L to decrease in such an amount that balance conditions are restored again.

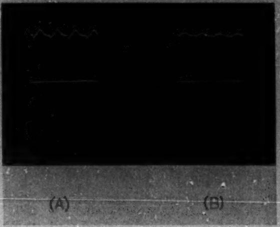


Fig. 2. Oscillograms of the waveform of the output voltage E_L of the constant-voltage regulator of Fig. 1

A—Without smoothing capacitor C_E
B—With a 4,000-microfarad capacitor C_E

A breadboard model of a 5-watt regulator corresponding to the circuit diagram of Fig. 1 has been investigated with regard to waveform of the output voltage, influence of load-current variations on actual magnitude of output voltage, and transient response.

The oscillograms of Fig. 2 show the waveform of output voltage E_L with full output for the following operating conditions: 1. The 4,000-microfarad electrolytic capacitor C_E is disconnected, and 2. this capacitor is parallel-connected to the output terminals, as shown in Fig. 1, to remove the remaining sharp-pulse a-c components of E_L. The ratio between peak value and average value of E_L is dependent on the magnitude of load current (0–4 amperes) and varies within the range of from 1.05–1.1.

In order to determine the influence of variations of load current I_L on actual magnitude of load voltage E_L, a variable resistor (0–100 ohms) has been connected in series with a moving-coil ammeter across the output terminals of the regulator; and a high-precision d-c potentiometer circuit has been employed for measuring the very small changes in magnitude of the output voltage E_L. The perform-

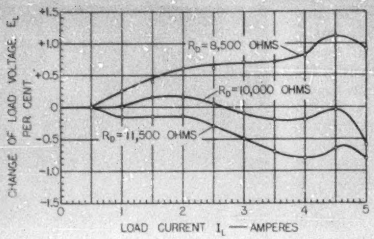


Fig. 3. Results of measurements on the constant-voltage regulator of Fig. 1: percentage change of output voltage E_L as a function of output current I_L with regeneration-control resistance R_D as a parameter

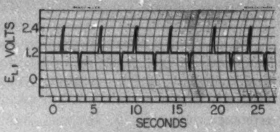


Fig. 4. Brush-recorder ink oscillogram characterizing the transient response of the constant-voltage regulator of Fig. 1

ance characteristics of Fig. 3 show the percentage change of E_L as a function of I_L with regeneration-control resistance R_D (Fig. 1) as a parameter. Evidently, with optimum adjustment (R_D=10,000 ohms), output voltage E_L varies within the limits of about ±0.2% in the entire range of load current I_L.

The transient response of the regulator has been measured by means of a magnetic-amplifier-operated ink oscillograph (Brush recorder).[9] The voltage (E_L) to be investigated is of the order of 1 volt, whereas the operating voltage of the Brush recorder is about 20 volts for full-scale deflection of the ink pen. Thus, the voltage gain (about 20) of the self-balancing magnetic amplifier operating the recorder has been adjusted so that E_L=1.0 volt corresponds to 80% of full-scale deflection (40 scale divisions) of the recorder.

Transient performance of the regulator has been tested by means of manually operated switch which connects and disconnects alternately a 0.3-ohm load resistor across the output terminals (output current I_L=4 amperes or zero, respectively). The test procedure consists in closing and opening the switch six times during a period of about 25 seconds. The corresponding oscillogram, Fig. 4, shows that, when closing the switch (I_L changes suddenly from 0–4 amperes), the output voltage E_L at first drops from its rated value of 1.2 volts to 0.4 volt and then, within about 0.5 second, reassumes the rated value. On the other hand, when opening the switch (I_L changes suddenly from 4 amperes to zero), E_L at first jumps from 1.2 volts up to 2.6 volts and then, within about 0.5 second, reassumes its rated value. In any case, after a small overshoot, the balance condition is restored. It is to be noted that, in the present problem (charging of storage batteries), the transient performance of the regulator has minor importance, because the load current I_L changes slowly during the charging procedure.

Referring to the performance characteristics of the simple constant-voltage reference circuit, as shown in Fig. 1, it is to be noted that in the specific problem outlined in a previous section of this paper, a regulation accuracy of ±2% was required at room temperatures for ±10% variations of the frequency-regulated 60-cycle power supply. With these moderate requirements, the accuracy of regulation of this form of constant-voltage reference circuit (about ±1% or better) has proved to be sufficient, and application of some frequency-compensating circuit has been considered not to be necessary in this case.

Constant-Light-Intensity Regulator for Instrumentation

General

Saturable-reactor-type constant-voltage regulators have been used successfully for eliminating those changes in light intensity of metal-filament lamps, which are caused by voltage fluctuations of the a-c power supply. This method of regulation, of course, involves the basic assumption that the performance characteristics of the lamp to be regulated, i.e., the properties of its metal filament, remain unchanged over long periods of time. Actually, however, as a result of "aging" effects, the relationship between applied voltage and light intensity changes considerably when operating a metal filament lamp with its rated value of voltage.

An improved method for constant-light-intensity regulation consists in applying a magnetic-amplifier regulator in conjunction with a barrier-layer-type photoelectric cell in such a way that the regulation is inherently independent of aging of the lamp.

In this case, the automatic regulator (besides its usual performance) takes care of the changes in performance characteristics of the lamp. That is, within certain limits, the regulator, automatically and continuously, also eliminates those effects, which are caused by changes of the relationship between light intensity and applied voltage or applied current, respectively.

The Specific Problem

In an actual case, the problem consisted of providing an area of constant light intensity (constancy: ±1%) for instrumentation. A conventional-type low-voltage lamp (rated voltage 5 volts, corresponding filament current about 2 amperes) is to be used in such a way that changes in performance of the lamp caused by aging will have no appreciable effect on the ±1% accuracy of regulation.

The general approach was to employ a magnetic-amplifier regulator in conjunction with a barrrier-layer-type photoelectric cell in such a way that the output current of this cell is constant within the limits of ±1%. It was considered to be permissible to apply here the basic assumption that the performance characteristics of the barrier-layer-type photoelectric cell remain unchanged, i.e., the relationship between light intensity and photoelectric output current is constant within the limits of ±1% over long periods of time.

The essential part of this specific problem was characterized as follows. If a light-absorbing body, for example, a plate of glass or some similar material producing an absorption of about 10–20%, is placed between the lamp and the light-sensitive surface of the photoelectric cell, then the magnetic-amplifier regulator will automatically compensate for this absorption so that the output current of the cell remains substantially unchanged. This means that in this case, the regulator must increase the voltage applied to the lamp exactly in such an amount that the increased light intensity of the lamp does compensate for the loss of radiation produced by the light-absorbing body, which has been placed between the lamp and the cell.

Application of a self-balancing magnetic amplifier in conjunction with a high-performance barrier-layer-type photoelectric cell made a simple solution of this problem possible. This novel approach is based upon the recognition that the combination of an a-c-supplied metal-filament lamp

Vol.
Pt. 1
1959

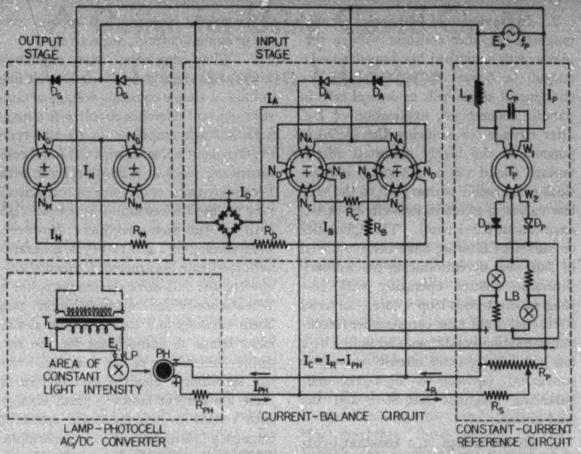


Fig. 5. Circuit diagram of the constant-light-intensity regulator with a self-balancing magnetic amplifier operating under current-balance conditions

with such a d-c producing photoelectric cell is a special form of an a-c/d-c converter, which lends itself particularly well to application with the self-balancing magnetic-amplifier technique.

In accordance with this concept, the "lamp-photocell a-c/d-c converter" is introduced into the negative-feedback channel of a self-balancing magnetic-amplifier circuit with infinite internal gain operating under current-balance conditions. This approach offers various advantages concerning performance characteristics of a constant-light-intensity regulator which automatically and continuously compensates for the effects produced by aging of the lamp to be regulated.

The Regulator Circuit

Fig. 5 shows a circuit diagram of the constant-light-intensity regulator, which has four main parts: 1. The constant-current reference circuit providing the constant reference current I_R. 2. The lamp photocell a-c/d-c converter producing a direct output current I_{PH} which is a measure for the actual light intensity influencing the photocell. 3. A self-balancing magnetic amplifier (input stage) with the current-balance circuit (resultant control current $I_C = I_R - I_{PH}$) and a full-wave rectifier in the output circuit. 4. A self-saturating magnetic amplifier (second stage) controlled by the full-wave-rectified output current I_M of the input stage and supplying a variable alternating output current I_N to the lamp-photocell a-c/d-c converter.

Because this regulator may sometimes be supplied from a small 60-cycle generator with moderate accuracy of frequency regulation, a special form of two-stage constant-current reference circuit has been employed in this case. The first stage, a magnetic voltage stabilizer (similar to that of Fig. 1) eliminates the disturbing effects of sudden changes in magnitude of power-supply voltage E_P and supplies the full-wave rectifier circuit with two silicon diodes D_P. The output voltage of this rectifier circuit, however, is linearily proportional to power-supply frequency f_P. In order to compensate for disturbing effects of variations of f_P, the constant-current reference circuit has a second stage, which consists of a Wheatstone-bridge network *LB* with two linear resistors and two nonlinear resistors (small metal-filament lamps, 6 volts, General Electric type *46*). The thermal inertia of these lamps is small enough with regard to the moderate speed of frequency variations of the a-c generator.

Additionally, the lamp-bridge network *LB* compensates for changes in voltage drops across rectifiers D_P and also for the effect of variations in ambient temperature on the saturation-flux density of the Orthonol core of the magnetic voltage stabilizer. Thus, in the 2-stage reference circuit of Fig. 5, the first stage takes care of sudden changes in magnitude of power-supply voltage, whereas the second stage performs the following three functions: 1. it compensates for the lower-speed frequency variations of the a-c generator, 2. it eliminates the effect of variations of voltage drops across the full-wave rectifier elements, and 3. it eliminates the effect of ambient temperature on the saturation-flux density of the saturating-core reactor of the magnetic voltage stabilizer of the first stage.

The self-balancing magnetic-amplifier circuit (input stage) comparing actual magnitudes of the control-current components I_R and I_{PH} is similar to that of Fig. 1. However, because the input-power level here is much smaller than that of the arrangement of Fig. 1, Supermalloy 2-mil tape cores have been used in this case.

The second stage consisting of a conventional-type self-saturating circuit with Orthonol 2-mil tape cores, gate windings N_G, d-c control windings N_M, and control-circuit series resistor R_M supplies the variable alternating current I_N to the primary winding of output transformer T_L.

The lamp-photocell a-c/d-c converter contains a 5-volt 2-ampere lamp (General Electric type *1709*) connected across the secondary winding of the step-down transformer T_L (voltage ratio 50/5 volts), and a high-performance barrier-layer-type photocell PH (Weston "Photronic Photoelectric Cell", model *856*, hermetically sealed) with series resistor R_{PH} in its output circuit.

Principle of Operation and Performance Characteristics

When placing a plate of glass or similar material producing an absorption of about 10–20% between the lamp and the light-sensitive surface of the photoelectric cell, the intensity of radiation influencing this cell is reduced. As a result, during the transient operation of the regulator, the output current I_{PH} of the cell drops in a certain amount, i.e., $I_{PH} < I_R$, and (because I_R is constant) a resultant control current $I_C = I_R - I_{PH}$ of positive direction causes alternating output current I_A of the first stage and the corresponding control current I_M of the second stage to decrease. Thus, output current I_N, lamp current I_L, lamp voltage E_L, and the light intensity influencing the photocell will increase in such an amount that balance conditions are restored. On the other hand, if the light-absorbing body is removed, then $I_{PH} > I_R$, and a resultant negative control current $I_C = I_R - I_{PH}$ will cause increasing of currents I_A, I_M, and corresponding decreasing of I_L, E_L and light intensity in such an amount that balance conditions are restored again.

A breadboard model of the constant-light-intensity regulator corresponding to the circuit diagram of Fig. 5 has been investigated with regard to accuracy of regulation and transient response, which is influenced by the properties of the regulator circuit itself and also by the thermal inertia of the lamp.

In order to simulate changes in performance characteristics of the lamp, caused by the effect of aging over longer periods of time, transparent plates of various thicknesses have been introduced between the lamp and the light-sensitive surface of the photoelectric cell. The amount of absorption varied actually in the range of about 10–20%. Larger rates of absorption have been produced by partially covering the cell by a metal plate. In any case, the magnitude of output current I_{PH} of the cell remained constant, and the accuracy of regulation was ±1% or better, when using infinite internal gain of the self-balancing magnetic-amplifier circuit. The associated variations of the rms value of lamp voltage E_L (measured with an iron-vane voltmeter) are in the range of from about 4.8–5.6 volts.

The transient performance of this regulator has been investigated by suddenly removing an absorption plate, which covers the light-sensitive surface of the cell. In this case, after a transient operation of 0.6 second balance conditions of the regulator circuit are restored. If larger variations of power-supply voltage E_P are expected, particularly with sudden changes of E_P, it is advisable to supply the regulator circuit from a conventional-type magnetic voltage stabilizer.

Conclusions

1. Application of self-balancing magnetic amplifiers with infinite internal gain ("critical regeneration") in automatic regulators, as described in this paper, offers various advantages concerning the performance characteristics of such regulators. However, it is to be noted that these amplifiers operate in principle under natural magnetization conditions. Therefore, the output-voltage waveshape corresponding to that of conventional forms of high-gain self-saturating circuits is highly distorted. In many cases it is possible to apply *L-C* filter circuits in the usual way for reducing the a-c components of the unidirectional output voltage of the amplifier to sufficiently small values.

2. In the design of a 60-cycle-operated magnetic-amplifier constant-voltage regulator with smoothed low-impedance d-c output (a 5-watt battery charger for charging of single cells or several cells in parallel connection), application of *L-C* filter circuits for reducing the a-c components of the highly distorted output voltage (rated value 1.12 volts) is practically impossible with regard to the very low values of effective load-circuit resistance (about 0.2–1 ohm). The difficulties on sufficient filtering have been removed in this case by introducing an ordinary saturable reactor operating with rectangular-hysteresis-loop core material (Orthonol 2-mil tape cores) under forced-magnetization conditions and acting as a low-gain square-wave circuit (waveform converter) between a high-gain self-balancing magnetic amplifier and the load circuit.

3. In the design of a constant-light-intensity regulator for instrumentation, a novel approach consists in applying a magnetic-amplifier regulator in conjunction with a high-performance barrier-layer-type photoelectric cell in such a way that the regulation is inherently independent of aging of the metal-filament lamp to be regulated. In this case, besides its usual performance, the regulator takes care of the changes in performance characteristics of the lamp. That is, within certain limits, the regulator automatically and continuously eliminates those disturbing effects, which are caused by changes of the relationship between light intensity and applied voltage or applied current, respectively. This approach, however, involves the basic assumption that the relationship between light intensity and output current of the photoelectric cell is constant within the limits of ±1% over long periods of time.

4. The transient performance of both regulators, as described in this paper, may be described by a time constant of about 0.4–0.6 second. Although the principle of operation of these regulators includes elimination of the effects of changes in magnitude of the 60-cycle power-supply voltage, it may be advisable in some cases to supply the regulator circuit from a conventional-type magnetic voltage stabilizer to eliminate the effects of sudden larger changes of the power-supply voltage. It is to be noted that the main purpose of these regulators is to eliminate the effects of changes arising in the regulator circuit itself, e.g., variable voltage drops across transformer windings and rectifier elements, and changing of the performance of metal-filament lamps as a result of aging, i.e., changes of the relationship between light intensity and applied voltage or applied current, respectively.

5. When supplying these regulators from the frequency-regulated 60-cycle line voltage, a simple constant-voltage or constant-current reference circuit with a magnetic voltage stabilizer (a parallel-type ferroresonant network with Orthonol-core saturating reactor, a linear reactor with Permalloy-dust core, and a capacitor) may be used successfully in conjunction with a full-wave rectifier. However, when employing a small 60-cycle generator with moderate accuracy of frequency regulation, the full-wave rectifier supplies a Wheatstone-bridge network with two linear resistors and two small metal-filament lamps to compensate for the frequency sensitivity of the magnetic voltage stabilizer. The thermal inertia of these low-current lamps is small enough with regard to the moderate speed of frequency variations of a-c generators. Additionally, the lamp-bridge network compensates for changes of voltage drops across the full-wave rectifier elements. Thus, in this simple form of a two-stage constant-voltage or constant-current reference circuit, the first (magnetic-core) stage takes care of sudden changes in magnitude of the power-supply voltage, while the second (lamp-bridge) stage compensates for the low-speed frequency variations of the a-c power supply and also eliminates the effects of variations of voltage drops across the full-wave rectifier elements and the effect of ambient temperature on the saturation-flux density of the Orthonol core of the magnetic voltage stabilizer.

References

1. Magnetic Amplifiers of the Self-Balancing Potentiometer Type, W. A. Geyger. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 71, 1952 (Jan. 1953 section), pp. 383–95.

2. Self-Balancing Magnetic Amplifier, W. A. Geyger. *U. S. Patent no. 2,700,130*, Jan. 18, 1955.

3. The Effective Feedback Factor of Self-Balancing Magnetic Amplifiers, W. A. Geyger. *AIEE Transactions*, pt. I, vol. 77, Mar. 1958, pp. 66–69.

4. The Transductor Amplifier (book), U. H. Krabbe. Lindhska Boktryckeriet, Orebro, Sweden, 1947, pp. 12–13, 17–20.

5. Avostat Controlled Dry Plate Rectifiers with Battery Reserve, A. U. Lamm. *ASEA Journal*, Allmänna Svenska Elektriska Aktiebolaget, Vasteras, Sweden, vol. 20, 1947, p. 3.

6. Transductors and Magnetic Amplifiers (book), A. G. Milnes. MacMillan and Company, Ltd., London, England; also St. Martin's Press, New York, N. Y., 1957, pp. 216–18.

7. Magnetic-Amplifier Circuits (book), W. A. Geyger. McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1957, pp. 46, 57, 282–83.

8. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 64, 73–74, 309.

9. Magnetic-Amplifier-Operated Ink Recorders, W. A. Geyger. *AIEE Transactions*, pt. I, vol. 77, Sept. 1958, pp. 457–71.

A Design for Using Closed-Circuit Television in Education

FRANCIS E. ALMSTEAD
NONMEMBER AIEE

THE TELEVISION medium cannot be used to solve all problems in the field of education. Its value as a means of communication for sight and sound will be as good as the quality of imagination and management which directs its utilization. Those who have professional courage can use its strength to improve and extend education. An increase in school population, increase in school costs and a growing shortage of experienced teachers are some of the rapid changes taking place in the field of education today. We all recognize that satisfactory solutions to each of these problems must be found in order to provide the quality education for our children which our society demands.

The American public has made considerable progress by building more schools and increasing teachers' salaries, but there is more to be done in providing equal educational opportunities for all youth and in upgradng the quality of education generally. School boards and professional educators are looking for new methods and devices which will help teachers meet their responsibility and which will create better learning conditions. With the rapid growth of technology during the past decade, it is natural for them to turn to electronics.

The Regents of the University of the State of New York recognized in electronic devices a potential for education and, in 1951, began to encourage the investigation and the use of television in education. Today, the New York State Education Department, under James E. Allen, Jr., Commissioner of Education, is conducting a very extensive investigation in all phases of this medium. Its four experiments in broadcast and closed-circuit television embrace many techniques and technical arrangements. One, a broadcast project, began on September 22, 1958, over channel 11 in New York City. These broadcasts can be viewed in four states, 8 hours a day, Monday through Friday. The viewing area in New York State has a million and a half public school children. In the Regents' broadcasting schedule there are six elementary grade subjects and six secondary school courses. Each of these follows a regular approved State syllabus.

Two other experiments are located in State University Teachers Colleges in Brockport and Albany where closed-circuit facilities are used from grade three through college. These systems use video distribution and a separate 2-way audio circuit. The facilities are within a single building at each college.

In Cortland County, the department is experimenting with closed-circuit facilities involving three school districts. Schools are connected by coaxial cable and use an independent pupil talkback system. In four locations, these facilities are financed by State appropriations and give the department the greatest undertaking ever attempted in educational communications by one institution.

The story of this latter system is told in three papers. This paper introduces the general aspects of the system and its basic educational requirements. A paper by Wall will discuss the technical aspects of the coaxial cable system.[1] A paper by Norton and Davidson will describe the design and operation of the independent talkback system which has never been attempted before on such a large scale.[2]

Educational Objective

In the rolling and hilly land of Cortland County, New York, educational television connects three school systems, one in a city and the other two in villages (Fig. 1). These are directed by two superintendents and three Boards of Education. Involved are the city of Cortland and the villages of Truxton and Virgil. Truxton is 9 miles and Virgil is 6 miles from the city of Cortland. Dr. Franklyn Barry, Superintendent of Schools in the city of Cortland, and Chester Lacey, District Superintendent, working in close cooperation with faculty members, proposed a plan whereby the three school systems could exchange benefits with each other through closed-circuit television. This plan incorporated four basic questions on the use of television facilities:

1. Can pupils in small schools have the advantages of services usually found only in large schools?
2. Can teaching talent be extended successfully into more than one classroom at a time?
3. Can the level of learning be raised?
4. Can schools provide courses which are now prohibited by excessive cost?

In the city of Cortland, there are about 4,500 pupils housed in five elementary schools and one junior-senior high school. In Truxton Central School there are 600 pupils in classes from kindergarten through the 12th grade. In Virgil Central School there are 400 pupils in these grades. By committee action it was decided to equip all second and fifth grade rooms in the elementary schools and a number of high school rooms in the three systems with television receivers. It was further decided to teach science to elementary school grades and basic art, English, health and mathematics to high school pupils. Teaching schedules began on September 29, 1958.

System Design Requirements

This system offers great promise and already has aroused much interest because of its uniqueness. Its electronic circuitry was planned through joint action by Dage (a Division of Thompson Products), Master Video Systems, the New York Telephone Company, and the New York State Education Department. To serve as headquarters for the project, and to house the studios, the city of Cortland Board of Education released a 2-room elementary school building, no longer in use (Fig. 2). The Board has refurbished the building. While origination may take place at any of the six Cortland schools, teaching will be done in the modern studios at the erstwhile "little red school house." Studios, control rooms, and classrooms have been equipped with electronic and television items by the State Education Department. Radio-frequency distribution and independent talkback systems have been installed by the New York Telephone Company and are being rented by the Department. The total system was planned with specific operating conditions in mind, some of which are:

1. Pupils should be able to see in full detail all of the intelligence transmitted by the television cameras without distraction or distortion.

Paper 59-194, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted October 20, 1958; made available for printing December 23, 1958.

FRANCIS E. ALMSTEAD is with the New York State Education Department, Albany, N. Y.

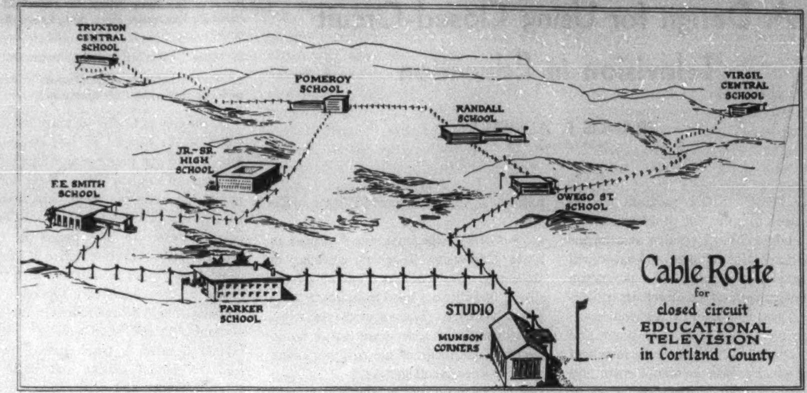


Fig. 1. System layout

2. Camera chains should be capable of faithfully reproducing the full intelligence in a picture, microscopic slide, printed matter, movie or transparency.

3. Natural atmosphere of a classroom should be retained in the viewing room.

4. Pupils should be able to hear all sounds without distortion, distraction, discomfort, or undue effort.

5. Pupils should be able to ask questions and receive immediate answers from the studio teacher.

6. System should be designed as an educational communications center capable of serving the entire school community with live lessons, motion pictures, slides, special programs and demonstrations, stereophonic recordings and 2-way audio programs.

7. System should permit separate transmissions to elementary and secondary schools simultaneously.

To meet these objectives two fully equipped studios and two radio-frequency channels have been provided. In addition, adequate distribution systems within each school, including high quality receivers, have been installed.

Studio Design and Equipment

Two studios are located in the converted 2-room school house. Design follows that used for commercial studios. Windows have been sealed, the walls have been treated with fiber glass held in place by chicken wire and covered with draped monk's cloth, a grid work of pipes for mounting lights has been hung from the ceiling, and independent electrical outlets have been placed systematically in the ceiling. Each outlet is controlled by a switch in the studio's control room. Lights can be operated independently, and the intensity of each light can be controlled by a dimmer. Thus, it is possible to obtain balanced light for any setting. A green chalk board with pin up boards on each end has been mounted on the wall facing the cameras. In order to obtain backlighting on the teacher and to reduce reflections from the board, it is held out from the wall with its upper edge extending forward from the vertical.

One of the studios is equipped with two professional Vidicon camera chains (Fig. 3). These are capable of fine quality performance with 600-line definition. The other studio is equipped with two other Vidicon camera chains which are capable of 500-line definition (Fig. 4). Two kinds of cameras were installed so that relative operational characteristics could be studied.

There is also a film chain suitable for showing 16-millimeter motion pictures and 35-millimeter slides and filmstrips (Fig. 5). It can be run independently or it can be used to support a teacher's presentation from the studio.

Chains operate through a console in the studio's control room (Fig. 6). From here the video and audio signals are transmitted to separate amplifiers in the

Fig. 2 (left). Studio building

Fig. 3 (right). A 600-line Vidicon camera

Fig. 4 (left). A 500-line Vidicon camera

Fig. 5 (right). Film chain

basement. Output from each video amplifier is a 1-volt, peak-to-peak, standard television signal. Quality control is essential, for unless the camera chains and audio equipment are operating at peak performance, it is impossible for the receivers to reproduce all of the intelligence the teacher hopes to convey. It is important to know the quality of the signal from the camera as well as the output from the video amplifiers. Thus, two *A* scopes (Fig. 7) are used in the control rooms for comparing camera output with video amplifier output. Change in signal composition can be observed immediately. Again for quality operation of the system, it is essential that the radio-frequency carrier be properly modulated all the time. A video-limiter amplifier is used to prevent overmodulation.

A good quality audio signal at 0 volume units across 600 ohms is supplied to the transmitter from either studio through an automatic gain control amplifier. This signal is monitored in the control room.

Studio lighting depends on the characteristics of the television pick-up tube. The Vidicon, for instance, requires more light than the image orthicon, but its nearly constant gamma characteristic permits a picture with tonal qualities approximating good photographs. In order to obtain this quality the incident light in a studio should be at least 200 foot-candles. The iris opening of the lens should be at f2.8. If 400 foot-candles of incident light are used, the iris may be reduced to f4.0. In both studios at Cortland up to 500 foot-candles of incident light are available. Full illumination is seldom used because of the excessive heat generated. It is, nevertheless, necessary to have reserve light to illuminate large areas or objects of small reflection.

Coaxial Cable Route

The coaxial cable route in Cortland is about 9 miles, the extension to Virgil is 6 miles, and to Truxton, about 11 miles. Situations representing various distances were chosen so that a cost analysis based on mileage and pupils served could be made. It seems logical to believe that a long cable to serve few pupils may not be practical. We hope to determine from experience with this installation the number of pupils required per mile of cable for an economically feasible operation.

Classroom Equipment and Utilization

Classrooms are equipped with two 24-inch or 21-inch table model receivers (to be modified by the addition of a quality speaker placed in front and below the set). The system of distribution in each school is designed to supply a minimum of 1,000 microvolts to each receiver. Each room has three 5-inch speakers serving the talkback system (Fig. 8). These

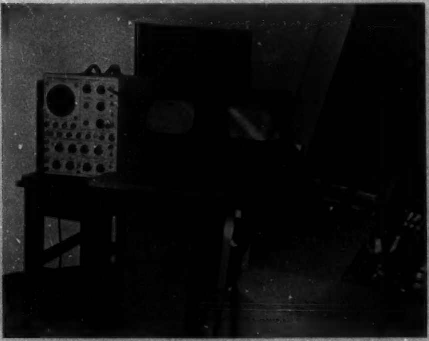


Fig. 6. Control room equipment

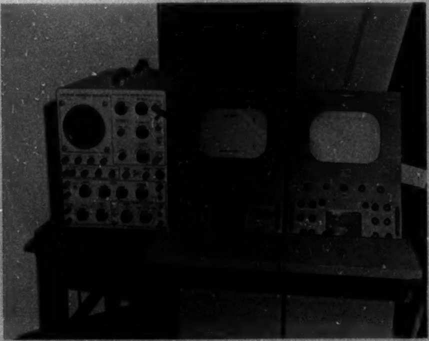


Fig. 7. Auxiliary "A" scope and monitors

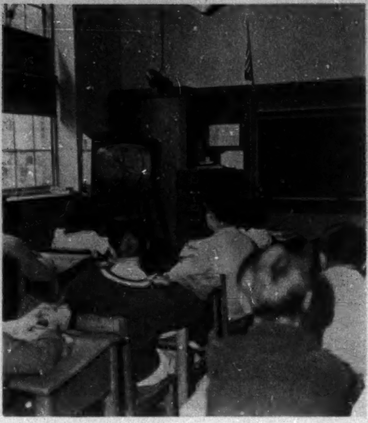


Fig. 8. Typical viewing and talkback equipment

speakers are used also as microphones and are suspended from the ceiling.

The talkback system is an important feature in the project. When used the pupil may talk to the teacher or the teacher may seek an answer from a pupil. Through the use of switches and signal lights the circuit can serve either the pupil or teacher. Pupils are not required to leave their seats when asking a question or giving an answer. Pupils in all other classrooms associated with the studio, as well as the studio teacher, will hear the child's voice clearly and without distortion. This system gives pupils the feeling of belonging and security through participation.

It is essential for all equipment from cameras in studios to receivers in classrooms to operate at peak performance at all times. Without excellent picture and sound quality throughout the system, transmission of intelligence will be substandard. This cannot be tolerated. It is, therefore, important for the project to have a full-time, well-qualified electronics technician. He has responsibility for the maintenance of all studio equipment and all television receivers at the schools. Experience may demonstrate the need for two technicians, one to concentrate on studio equipment, the other on receiver maintenance.

It is important that the televised lessons coming to the classrooms be of high quality, appropriate, and timely to learning (Fig. 9). It is of equal importance to create in the classrooms proper conditions and atmosphere (Fig. 10). Too many pupils viewing at one receiver, too great a distance between viewers and set, improper picture contrast and brightness, poor sound and poor picture definition will be deterrents to effective learning. Ideally, receivers, tipped slightly forward at the top, should be placed on stands in a room so that window glare does not interfere with the pupils' vision and so that the center of the picture is slightly above eye level of the viewing group (Fig. 8). It is also important for pupils to have those materials and supplies necessary to enable them to participate fully in the televised lesson. It is the responsibility of each teacher to make certain that classroom and receiver conditions are proper for maximum learning (Fig. 11).

Television Receivers

A television receiver when placed in a classroom has a role different from that which it plays in the home. A home viewing is casual and generally done for entertainment, while in the classroom it is the means of communicating intelligence to viewing pupils. Therefore, the circuitry of the receiver must reproduce the input message with a minimum of degradation. Design and performance of the set must be of high quality.

Experience has found the following technical considerations to be of great importance:

1. Ninety-degree picture tube, which gives better focus and less distortion than the 110-degree tube.
2. High-voltage power supply with an output of 18 kv, in no case less than 16 kv.
3. Variation in voltage greater than 2 kv between maximum and minimum brightness control adjustment.
4. Isolated chassis and a transformer power supply with heaters in parallel.
5. Adjustable keyed automatic gain control.
6. In the intermediate-frequency video section a band width of 4.5 megacycles and

Vol
Pt.
19

Fig. 9. Teacher in studio

Fig. 10. Typical classroom arrangement

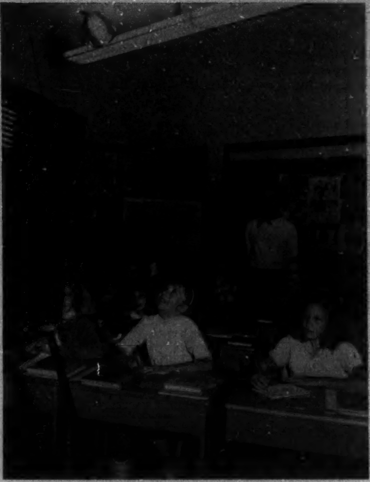


Fig. 11. Teacher in classroom

in the video amplifier a band width of 3.5 megacycles. These conditions make possible 300-line definition in the picture readable from a standard test pattern.

7. Brightness control range from black to 60 volts above cutoff.

8. Sensitivity range for lower channels at least 20 microvolts and for higher channels at least 40 microvolts to produce a 20-volt peak-to-peak input to the picture tube.

9. Audio amplifier with 5 watts of undistorted output.

10. Full range response speaker, 80 to 13,000 cycles, flat to ±3 db, baffled properly and mounted at the front.

Methods of Utilization

SIMULTANEOUS INSTRUCTION OF SMALL GROUPS

In communities where cost of education has become a burden, it may be impossible for boards of education to employ teachers for specialized work such as teaching of foreign languages, music, art, and science in elementary schools. A closed-circuit cable system like that installed in Cortland County provides a means by which boards of education may join together to employ an outstandingly well qualified person to teach each subject. A program for shared educational services over a cable system is working very well in Cortland county. Science taught by a studio teacher is available to all fifth grade pupils in the three school districts. Music is available this year for the first time to seventh and eighth grade pupils at Truxton and Virgil and is taught by a music teacher from the city of Cortland Board of Education.

USE BY INEXPERIENCED TEACHERS

In the upper grades particularly, inexperienced teachers need help in developing excellence of performance and competency in subject matter. While a new teacher is growing professionally, some pupils may be handicapped through inadequate teaching. This damage to learning can be avoided by the use of a television cable system which makes it possible for an experienced teacher to present authentic subject matter and skillful demonstrations to pupils wherever they are at one time. The teacher in the viewing room is responsible for pupil success. He has an opportunity to work with his pupils during two periods a week. In the case of high school physics, for example, the classroom teacher may use one of these periods for laboratory exercises and the other for drill, problem solving and examinations.

This team relation assures excellence of performance for all pupils and offers an opportunity to a new teacher for professional growth.

SPECIAL SUBJECT MATTER

In almost every school there are a few pupils who have exceptional ability. Schools, especially the smaller ones, are not able financially to employ a diversified staff to satisfy the needs of their gifted pupils. In several schools connected by cable there will be a sufficient number of superior boys and girls to justify the employment of proper faculty to serve the group.

TOTAL TEACHING

In certain situations, pupils in small groups can be taught effectively by a studio teacher having complete charge of the conduct of the course. Viewing rooms need not be supervised by qualified subject matter teachers. In some cases, teacher aides have been assigned to viewing rooms and in others a classroom chairman has been found to be satisfactory. To make this system effective the studio teacher and pupils need the advantages of 2-way audio communication. The advantages of this system are clear: Pupils may ask to have a point clarified or the teacher may ask a pupil to recite. All pupils, wherever they are, hear questions and recitations. In those subjects where total teaching is practical there can be a saving of manpower.

TEACHER TRAINING

This year, the Cortland system is being used 1 hour, 3 days a week, for a training program designed to improve instruction. Teachers in eight schools in three districts are enrolled. Television is the most convenient and economical method for conducting this type of interdistrict training program. The benefits received by faculty members through this training experience can be significant indeed.

System Expansion

This system of communications has been designed with the idea of expansion in mind. While the coaxial cable now accommodates two channels, a total of six channels can be made available. Cable could be extended to several smaller towns and villages to serve more classes of elementary and secondary school pupils. Expansion within present schools is also possible.

Broadcast reception in the southwest area of New York State is poor, so poor that many of the communities use community antenna systems. It is possible to feed all of these systems from the Cortland project. This distribution would cover ten cities and several smaller communities. In fact, this project has been designed to serve education in so many ways, expansion is almost limitless.

Problems for the Future

There has been, during the past 2 years, a great deal of experimentation at all levels of education involving many forms of educational television. Generally, there have been as many philosophies and objections as there have been centers of experimentation. Out of it all has come two observations: (1) There is no damage to learning when the medium of television and excellence of performance are combined to communicate a lesson and (2) when conditions are right pupils have accepted the use of the medium. As a matter of fact, there is evidence indicating pupils prefer television lessons when faced with a choice between having an outstanding teacher or not. Therefore, the questions of learning and acceptance by pupils have been resolved fairly well.

Assuming that television is an effective means of communication in education, then there are three other factors which need attention immediately:

1. How can the typical school schedule be altered so that television facilities can be used in making a maximum contribution for the learner?
2. How can the school curriculum be altered so that those facts and concepts most adaptable can be used on television?
3. How can professional educators, the classroom teachers, be encouraged to use the medium to benefit learners?

These are challenges which cannot be dealt with lightly. Administrators are interested in knowing the conditions which warrant a cable installation. One measure may be the number of pupils required per mile of cable so that the cost of pupil is reasonable.

Parents are requiring top quality education for their children. They are aware of the problems involved in providing quality and excellence in education with the pending teacher shortage, inexperienced teachers, and the mounting costs. They are looking to the use of television for answers to some of the growing problems. Therefore, it is important for engineers, technicians, and educators to join in a thoughtful, professional search to discover the most appropriate and effective use of television for the improvement of instruction in education.

References

1. A Transmission Design for Closed-Circuit Educational Television, W. L. Wall. *AIEE Transactions*, see pages 340–46 of this issue.

2. A Talkback System for Educational Television, R. C. Norton, J. B. Davidson. *Ibid.*, May 1959, pp. 209–14.

Vol.
Pt.
195

The Characteristic Impedance and Phase Velocity of a Shielded Helical Transmission Line

HERBERT S. KIRSCHBAUM
MEMBER AIEE

Synopsis: This paper presents the results of a theoretical and experimental investigation of the properties of a shielded helix when used as a transmission line. Since such a structure is dispersive and because of the presence of both TM and TE modes it cannot be said to have a characteristic impedance in the true sense. However, such an impedance is defined for low frequencies in terms of a distributed capacitance and inductance. These parameters, the capacitance and the inductance are determined in the case of infinite wave length for round wire helices. The equations are developed for a uniform dielectric inside the shield and for two dielectrics, one inside the helix and the other between the helix and the shield. The analytic results are then shown to agree very well with the experimental results obtained by the author.

HELICAL STRUCTURES have long been used in various ways as electromagnetic devices. In particular, they occur in the windings of core type transformers, video delay lines, traveling wave tubes, and as antenna structures. As a result of the interest in this structure, numerous reports have appeared describing the efforts made to analyze and measure the properties of a helix used as an electromagnetic device.

A helical structure can take many forms. Some of these are: a helix of infinite extent existing in an unbounded uniform medium; a helix of infinite extent in which the medium inside the helix radius is different from that outside; a shielded helix with either a uniform medium inside the shield or as in the previous situation, two media present. The greatest effort has been confined, in the past, to the first one of these forms, namely, the infinite helix in an unbounded uniform medium. The problem, in this case, is one of finding those solutions of the homogeneous Maxwell equations which satisfy the boundary conditions at the helix and at infinity. From these solutions, the various modes of operation of the helix can be evaluated.

In the case of a shielded helical transmission line, the problem is one of finding those solutions of the homogeneous Maxwell equations which satisfy the boundary conditions at the helix and at the shield. When used as a transmission line, the greatest interest is centered upon the characteristic impedance and phase constant of the helical line. Strictly speaking, the impedance concept is not a correct one when applied to a helical transmission line, since the waves present are a mixture of both transverse magnetic and transverse electric waves. Any impedances so developed apply only to relatively low frequencies, and they completely break down at a frequency where the circumference of the helix is equal to about one to two wave lengths. Beyond this frequency, the helix operates in many different modes, each with its own phase velocity.

This investigation treats the shielded helix as a distributed circuit transmission line. The inductance and capacitance parameters are derived for a quasi-round wire helix for infinite wave length. While these parameters have real meaning for direct current only, they should prove useful up to frequencies at which the wave length becomes equal to 1 to 2 circumferences. These inductance and capacitance parameters are used to derive expressions for the characteristic impedance and phase velocity of a shielded helix.

The Distributed Capacitance and Inductance

Because of the complexity involved in the development of the transmission line

---

Paper 59-266, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted August 22, 1958; made available for printing December 19, 1958.

Herbert S. Kirschbaum is with the Battelle Memorial Institute, Columbus, Ohio.

This paper is an abstract of a dissertation by this author which was presented to the Carnegie Institute of Technology, Pittsburgh, Pa., in partial fulfillment of the requirements for a degree of Doctor of Philosophy.

parameters, these derivations have been omitted. For those interested in the details of the derivations, reference 1 should be consulted. In this section, a brief description will be given of the methods used to develop the expressions for the transmission line parameters along with the final results. This will be followed by an application of these results to the problem of determining the characteristic impedance of the helical transmission line and also its phase velocity. They are then compared with experimental results obtained by this author.

Two of the parameters with which the behavior of a transmission line can be described are the distributed capacitance in, say, farads per meter of line length and the distributed inductance in, say, henrys per meter of line length. Since the transmission line of this report consists of a helix formed of round wire surrounded by a perfectly conducting sheath, there are two possible definitions of line length. The first of these might be measured in the axial direction. The second might be measured in the helix direction. Since the propagation of energy along such a line is actually in the axial direction, the former definition of line length is preferred. The capacitance and inductance parameters are developed per meter in the helix direction and then written in terms of line length in the axial direction.

The transmission line with which this paper is concerned consists of a helix of round wire formed by winding it upon a supporting tube or rod of dielectric material as shown in Figs. 1 and 2. The return conductor consists of a conducting sheath which is outside the helix and coaxial with the axis of the helix. The space between the helix and the sheath may be free space or may be filled with some dielectric such as oil, polystyrene, etc. For the purposes of determining the capacitance per unit length, the transmission line is assumed to be infinitely long.

In reference 1 a rather complete formulation of the problem of the capacitance is presented. It will suffice here to present some of the more salient features of the derivation. The method of approach used is to disregard the actual boundary at the surface of the wire and to replace it with a boundary that closely approximates the surface of the wire. The necessity for doing this lies in the intractability of Laplace's equation when written in a co-ordinate system which will describe the surface of the helical wire. Furthermore, any such co-ordinate system would not contain the sheath as one of its co-ordinate surfaces.

A potential function is derived which satisfies Laplace's equation in a circular cylindrical co-ordinate system and which also satisfies the boundary conditions, that the potential at the sheath is zero and that the discontinuity in the radial component of the displacement vector, at the helix radius, be equal to the surface charge density on a very narrow tape helix. When the width of this tape is allowed to approach zero, the tape helix degenerates into a line helix of infinitesimal cross section having a uniform charge of Q_l coulombs per meter in the helix direction. The potential function so derived approaches infinity as one approaches the line helix.

Considering only a region in the immediate vicinity of the line, the equipotential surfaces are very close to right circular cylinders formed into the shape of a helix. This is the case since, as the line charge is approached very closely, the helical line appears to approach an infinitely long straight line of charge and it is known that the equipotential surfaces around such a line of charge are right circular cylinders. One of these equipotential surfaces is chosen as the surface of the wire. Although such a surface does not coincide with the actual wire surface, the agreement between it and the wire for very small wires is sufficiently close for practical applications. The potential of this equipotential surface when divided by the charge in coulombs per meter yields the reciprocal of the distributed capacitance. The procedures outlined above are carried out in detail in reference 1, the results being

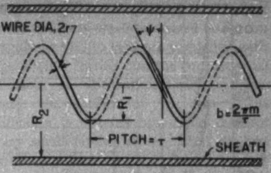


Fig. 1. Shielded helix showing dimensions

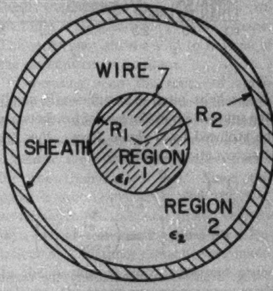


Fig. 2. Cross section of shielded helix showing regions 1 and 2

Any consistent system of length units may be used since all lengths enter as dimensionless ratios. The permitivities ϵ_1 and ϵ_2 are those of regions 1 and 2 defined in Fig. 2. Normally ϵ_1 is greater than ϵ_2, but in the case where they are equal, equation 2 reduces to

$$\frac{1}{C}=\frac{R_1}{\epsilon_2 r\cos\psi}\left\{\operatorname{Log}_e\frac{R_2}{R_1}+2\sum_{m=1}^{\infty}\frac{\left[1-\frac{I_m(bR_1)K_m(bR_2)}{I_m(bR_2)K_m(bR_1)}\right]I_m(bR_1)K_m(bR_1)\cos\left(\frac{br}{\cos\psi}\right)}{\frac{\epsilon_1}{\epsilon_2}+\left(\frac{\epsilon_1}{\epsilon_2}-1\right)\left[1-\frac{I_m'(bR_1)K_m(bR_2)}{I_m(bR_2)K_m'(bR_1)}\right](bR_1)I_m(bR_1)K_m'(bR_1)}\right\} \tag{1}$$

where b is equal to $(2\pi m/\tau)$. Or

$$\frac{1}{C_z}=\frac{1}{2\pi\epsilon_2}\left\{\operatorname{Log}_e\frac{R_2}{R_1}+2\sum_{m=1}^{\infty}\frac{\left[1-\frac{I_m(bR_1)K_m(bR_2)}{I_m(bR_2)K_m(bR_1)}\right]I_m(bR_1)K_m(bR_1)\cos\left(\frac{br}{\cos\psi}\right)}{\frac{\epsilon_1}{\epsilon_2}+\left(\frac{\epsilon_1}{\epsilon_2}-1\right)\left[1-\frac{I_m'(bR_1)K_m(bR_2)}{I_m(bR_2)K_m'(bR_1)}\right](bR_1)I_m(bR_1)K_m'(bR_1)}\right\} \tag{2}$$

where C is the capacitance per meter in the helix direction and C_z is the capacitance per meter in the axial direction. The I_m and K_m functions are modified Bessel functions of order m and of the first and second kinds, respectively. The I_m' and K_m' functions are derivatives of the modified Bessel functions with respect to the argument $(b\rho)$. The other dimensions are as defined in Figs. 1 and 2.

$$\frac{1}{C_z}=\frac{1}{2\pi\epsilon_2}\left\{\operatorname{Log}_e\frac{R_2}{R_1}+2\sum_{m=1}^{\infty}\left[1-\frac{I_m(bR_1)K_m(bR_2)}{I_m(bR_2)K_m(bR_1)}\right]\times I_m(bR_1)K_m(bR_1)\cos\left(\frac{br}{\cos\psi}\right)\right\} \tag{3}$$

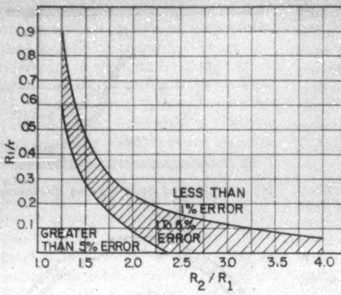


Fig. 3. Estimate of errors expected using simplified expressions for capacitance and inductance

If the helix pitch τ is sufficiently small, the summation in the pair of brackets can be replaced by a closed form. For this case, equation 2 becomes,

$$\frac{1}{C_z}=\frac{1}{2\pi\epsilon_2}\left\{\log_e \frac{R_2}{R_1}-\frac{2\epsilon_2}{\epsilon_1+\epsilon_2}\times \sin\psi\cdot\log_e\left[2\sin\left(\frac{\pi r}{\tau\cos\psi}\right)\right]\right\} \tag{4}$$

Fig. 3 shows approximately the range of R_1/τ for which equation 4 can be used and still be within a specified percentage of equation 2. This figure actually shows the error to be expected in the characteristic impedance resulting from the use of equation 4 and a similar expression for inductance. Since the capacitance, inductance, and impedance expressions are very similar, the curves in Fig. 3 give a good indication of the range of helix dimensions for which equation 4 is applicable.

If on the other hand the helix pitch τ approaches infinity, the inner conductor becomes a straight wire parallel to the axis and displaced from it by a distance R_1. Within the radius R_1 the permittivity is ϵ_1, and between R_1 and R_2 it is ϵ_2. For very long pitches the capacitance becomes approximately

$$\frac{1}{C_z}=\frac{1}{2\pi\epsilon_2}\left\{\log_e \frac{R_2}{R_1}+\frac{2\epsilon_2}{\epsilon_1+\epsilon_2}\times \sum_{m=1}^{\infty}\frac{\left[1-\left(\frac{R_1}{R_2}\right)^{2m}\right]\cos\left(\frac{mr}{R_1}\right)}{m\left[1-\left(\frac{\epsilon_1-\epsilon_2}{\epsilon_1+\epsilon_2}\right)\left(\frac{R_1}{R_2}\right)^{2m}\right]}\right\} \tag{5}$$

For the case in which $\epsilon_1=\epsilon_2$ this becomes

$$\frac{1}{C_z}=\frac{1}{2\pi\epsilon_2}\left\{\log_e \frac{R_2}{R_1}+ \sum_{m=1}^{\infty}\left[1-\left(\frac{R_1}{R_2}\right)^{2m}\right]\frac{\cos\left(\frac{mr}{R_1}\right)}{m}\right\} \tag{6}$$

This can be written in the following closed form:

$$\frac{1}{C_z}=\frac{1}{2\pi\epsilon_2}\left\{\log_e \frac{R_2}{R_1}+ \frac{1}{2}\log_e\left[\frac{1-2\left(\frac{R_1}{R_2}\right)^2\cos\frac{r}{R_1}+\left(\frac{R_1}{R_2}\right)^4}{2\left(1-\cos\frac{r}{R_1}\right)}\right]\right\} \tag{7}$$

For very small r/R_1 as compared with unity, this becomes

$$\frac{1}{C_z}=\frac{1}{2\pi\epsilon_2}\left\{\log_e \frac{R_2}{r}\left(1-\frac{R_1^2}{R_2^2}\right)\right\} \tag{8}$$

The exact expression for the capacitance between two parallel cylinders which are not coaxial is given in reference 2, page 76. The capacitance there derived, in the notation of this author, is

$$\frac{1}{C_z}=\frac{1}{2\pi\epsilon_2}\cosh^{-1}\left\{\frac{R_2}{2r}\left(1-\frac{R_1^2}{R_2^2}\right)+\frac{r}{2R_2}\right\} \tag{9}$$

This expression will approach equation 8 as (r/R_2) becomes very small. Depending upon the range of applicability, any one of the capacitances presented above may be used with the inductance per meter, to arrive at a characteristic impedance and phase velocity for the transmission line. Since this report is concerned mainly with characteristic impedance and phase velocity, no calculations of capacitance as such were made. In the sections on impedance and phase velocity, the above expressions were used, where applicable, to determine a family of impedance and phase velocity curves.

In reference 1 a rather complete formulation of the problem of the distributed inductance is given. As in the case of the capacitance, the actual wire is not used to form the boundary of the helix. Instead, a different surface which closely approximates the actual wire is used in place of the wire. It is shown that the inductance of the helix can be arrived at be a complete knowledge of the vector potential of electric current in the region surrounding the helix. The inductance so calculated neglects any internal flux in the wire, this assumption being based upon the fact that the frequencies at which this structure is used, the current in the wire will be confined to the surface of the wire. The vector potential is arrived at by solving in cylindrical coordinates what might be called the vector equivalent of Laplace's equation, along with the requirement that the divergence of the vector potential be zero. These two equations, which must be satisfied in the magnetostatic case, are

$$\nabla\times\nabla\times A=0 \tag{10}$$

$$\nabla\cdot A=0 \tag{11}$$

where A is the vector potential of electric current in the current free region surrounding the helix. The vector potential and the magnetic field which can be derived from it must satisfy certain boundary conditions. The most important of these is, that the magnetic field outside the sheath be zero and that the discontinuity in the tangential component of the H vector at the helix radius, R_1, be numerically equal to the current density in a very narrow tape helix at R_1. When the width of the tape is allowed to approach zero, the tape degenerates into a line helix of infinitesimal cross section carrying a current of I amperes. The vector potential so derived approaches infinity as one approaches the helical line. If one approaches the line sufficiently close, the distance between the line and the observer becomes small compared with the radius of curvature of the line. In this case the helical line may be approximated by an infinite straight line carrying current. The surfaces of equi-vector-potential which surround such a straight line of current are known to be right circular cylinders. A round wire carrying current in which the current density in the wire is a function of the radius only, also has equi-vector-potential surfaces which are right circular cylinders. In fact the wire surface itself is an equi-vector-potential surface. Therefore, as an approximation, one of the equi-vector-potential surfaces which surrounds the line helix is used as the surface of the wire. This approximation approaches perfection as the wire radius approaches zero. As is shown in reference 1, the magnitude of the vector potential in the helix direction at the wire surface when multiplied by (μ_0/I) yields the inductance per meter in the helix direction.

Vol.
Pt. 1
1959

The procedure outlined in the preceding paragraph is carried out in detail,[1] and the results of this work are given in several forms:

$$L=\frac{\mu_0 R_1 \cos\psi}{\tau}\left\{\frac{1}{2}\left(1-\frac{R_1^2}{R_2^2}\right)+\tan^2\psi\cdot\log_e\frac{R_2}{R_1}-2\sum_{m=1}^{\infty}\left[1-\frac{I_m'(bR_1)K_m'(bR_2)}{I_m'(bR_2)K_m'(bR_1)}\right]\times I_m'(bR_1)K_m'(bR_1)\cos\left(\frac{br}{\cos\psi}\right)\right\} \quad (12)$$

where L is the inductance in henrys per meter in the helix direction. The other dimensions are defined in Figs. 1 and 2. Any consistent system of length may be used since all lengths enter as dimensionless ratios. If L_z is used to denote the inductance per meter in the axial direction, equation 12 becomes

$$L_z=2\pi\mu_0\left(\frac{R_1}{\tau}\right)^2\left\{\frac{1}{2}\left(1-\frac{R_1^2}{R_2^2}\right)+\tan^2\psi\cdot\log_e\frac{R_2}{R_1}-2\sum_{m=1}^{\infty}\left[1-\frac{I_m'(bR_1)K_m'(bR_2)}{I_m'(bR_2)K_m'(bR_1)}\right]\times I_m'(bR_1)K_m'(bR_1)\cos\left(\frac{br}{\cos\psi}\right)\right\} \quad (13)$$

If the helix pitch is sufficiently small, the summation in the brackets can be replaced by a closed form. For this case equation 13 becomes

$$L_z=2\pi\mu_0\left(\frac{R_1}{\tau}\right)^2\left\{\frac{1}{2}\left(1-\frac{R_1^2}{R_2^2}\right)+\tan^2\psi\times\left[\log_e\frac{R_2}{R_1}-\frac{1}{\sin\psi}\log_e\left(2\sin\left(\frac{\pi r}{\tau\cos\psi}\right)\right)\right]\right\} \quad (14)$$

Fig. 3 shows the range of R_1/τ for which equation 14 can be used with reasonable accuracy. Although this figure is strictly applicable to the characteristic impedance, it is sufficiently accurate to indicate for what range of (R_1/τ) good results can be expected from equation 14.

For a very close wound helix of small wire, the second term within the bracket becomes negligible and the above expression reduces to,

$$L_z=\pi\mu_0\left(\frac{R_1}{\tau}\right)^2\left(1-\frac{R_1^2}{R_2^2}\right) \quad (15)$$

This checks previously derived results[2] for very close wound helices in shielding cans.

On the other hand, if the pitch of the helix becomes very large it approaches a straight wire parallel to the axis of the sheath and displaced from the axis by a distance of R_1. Hence for a very long pitch the inductance becomes approximately

$$L_z=\frac{\mu_0}{2\pi}\left\{\log_e\frac{R_2}{R_1}+\sum_{m=1}^{\infty}\left[1-\left(\frac{R_1}{R_2}\right)^{2m}\right]\frac{\cos\left(\frac{mr}{R_1}\right)}{m}\right\} \quad (16)$$

This can be expressed in the following closed form:

$$L_z=\frac{\mu_0}{2\pi}\left\{\log_e\frac{R_2}{R_1}+\frac{1}{2}\times\log_e\left[\frac{1-2\left(\frac{R_1}{R_2}\right)^2\cos\frac{r}{R_1}+\left(\frac{R_1}{R_2}\right)^4}{2\left(1-\cos\frac{r}{R_1}\right)}\right]\right\} \quad (17)$$

For very small r/R_1 as compared with unity, this becomes

$$L_z=\frac{\mu_0}{2\pi}\left\{\log_e\frac{R_2}{r}\left(1-\frac{R_1^2}{R_2^2}\right)\right\} \quad (18)$$

When this expression is used along with equation 8 to compute the characteristic impedance of this degenerate helix ($\tau=\infty$), there results an expression which is in agreement with the accepted[4] impedance for this structure. This validates equation 14 for the case of infinitely long pitch.

The expressions in this section for the inductance and in the previous section for the capacitance are based upon very small wire radii. However, if the pitch of the helix becomes of the same order of magnitude as the wire diameter, the wire radius, while still very small compared to R_1, can hardly be said to be small compared with the pitch. In this case the equipotential surfaces do not satisfactorily approximate the surface of the wire. This difficulty is vitiated by noting that for very small pitches all terms but the first within the braces of equation 13 become negligibly small. Since the first term is independent of the wire diameter it may be concluded that for very close wound small wire helices the distributed inductance is independent of the wire shape. The same may also be said to be true of the distributed capacitance.

Depending upon the range of applicability, any one of the above inductances may be used with the appropriately chosen capacitance to arrive at a characteristic impedance and phase velocity for the transmission line. Since this paper is concerned mainly with the characteristic impedance and phase velocity of a helical transmission line, no calculations of inductance as such were made. In the section on impedance and phase velocity, the above expressions were used, where applicable, to determine a family of impedance and phase velocity curves.

Characteristic Impedance and Phase Velocity

When a distributed circuit transmission line which has an inductance L_z per unit length and a capacitance C_z per unit length is studied, it is found that it can sustain traveling waves of current and voltage. In the absence of losses, it is found that these waves travel unattenuated with a velocity v_z which is independent of frequency. At any point the ratio of voltage to current is also a constant Z_0 which is independent of frequency. These two constants, the phase velocity and characteristic impedance, are related to the inductance and ca-

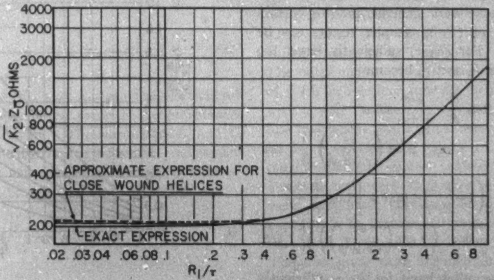


Fig. 4. Comparison of various expressions for characteristic impedance

pacitance parameters as follows:

$$Z_0=\sqrt{\frac{L_2}{C_2}} \tag{19}$$

$$v_z=\frac{1}{\sqrt{L_2C_2}} \tag{20}$$

The inductance and capacitance parameters were derived upon the assumption of infinite wave length. Consequently these constants have meaning only so long as the wave length is quite long compared with the pitch of the helix. The basis for this assertion is the fact that the mutual effect of the turns of the helix upon any one given turn becomes smaller as the distance between them increases. Therefore, it is the turns closest a given one that have the greatest effect upon its inductance and capacitance. If the wave length is long, the turns in the vicinity of any one given turn are essentially at the same potential and carry essentially the same current and the above parameters have real meaning.

By using equation 2 and 13 in equation 19 one obtains for the characteristic impedance,

$$Z_0=\frac{R_1}{\tau}\sqrt{\frac{\mu_0}{\epsilon_2}}\cdot\sqrt{F_1F_2} \tag{21}$$

where

$$F_1=\left\{\log_e\frac{R_2}{R_1}+2\sum_{m=1}^{\infty}\frac{\left[1-\frac{I_m(bR_1)K_m(bR_2)}{I_m(bR_2)K_m(bR_1)}\right]I_m(bR_1)K_m(bR_1)\cos\left(\frac{br}{\cos\psi}\right)}{\frac{\epsilon_1}{\epsilon_2}+\left(\frac{\epsilon_1}{\epsilon_2}-1\right)\left[1-\frac{I_m'(bR_1)K_m(bR_2)}{I_m(bR_2)K_m'(bR_1)}\right](bR_1)I_m(bR_1)K_m'(bR_1)}\right\} \tag{22}$$

$$F_2=\left\{\frac{1}{2}\left(1-\frac{R_1^2}{R_2^2}\right)+\tan^2\psi\cdot\mathrm{Log}_e\frac{R_2}{R_1}-2\sum_{m-1}^{\infty}\left[1-\frac{I_m'(bR_1)K_m'(bR_2)}{I_m'(bR_2)K_m'(bR_1)}\right]\times I_m'(bR_1)K_m'(bR_1)\cos\left(\frac{br}{\cos\psi}\right)\right\} \tag{23}$$

The expression given by equation 21 is of course very general and for certain ratios of (R_1/τ) simpler results can be used. For (R_1/τ) sufficiently large, the factors F_1 and F_2 become

$$F_1=\left\{\log_e\frac{R_2}{R_1}-\frac{2\epsilon_2}{\epsilon_1+\epsilon_2}\sin\psi\cdot\log_e\left[2\sin\left(\frac{\pi r}{\tau\cos\psi}\right)\right]\right\} \tag{24}$$

$$F_2=\left\{\frac{1}{2}\left(1-\frac{R_1^2}{R_2^2}\right)+\tan^2\psi\cdot\left[\mathrm{Log}_e\frac{R_2}{R_1}-\frac{1}{\sin\psi}\cdot\log_e\left(2\sin\left(\frac{\pi r}{\tau\cos\psi}\right)\right)\right]\right\} \tag{25}$$

On the other hand for R_1/τ sufficiently small, using equations 5 and 18,

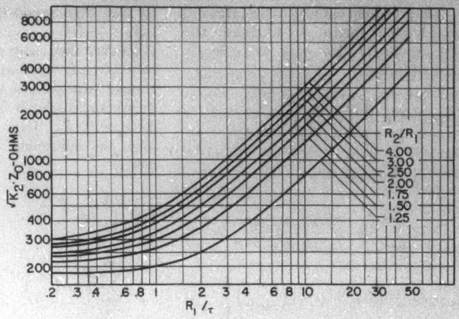


Fig. 5. Characteristic impedance versus R_1, τ, and R_2

$$Z_0=\frac{1}{2\pi}\sqrt{\frac{\mu_0}{\epsilon_2}}\cdot\sqrt{G_1G_2} \tag{26}$$

where

$$G_1=\left\{\log_e\frac{R_2}{R_1}+\frac{2\epsilon_2}{\epsilon_1+\epsilon_2}\times\sum_{m=1}^{\infty}\frac{\left[1-\left(\frac{R_1}{R_2}\right)^{2m}\right]\cos\left(\frac{mr}{R_1}\right)}{m\left[1-\left(\frac{\epsilon_1-\epsilon_2}{\epsilon_1+\epsilon_2}\right)\left(\frac{R_1}{R_2}\right)^{2m}\right]}\right\} \tag{27}$$

$$G_2=\left\{\log_e\frac{R_2}{r}\left(1-\frac{R_1^2}{R_2^2}\right)\right\} \tag{28}$$

By using equations 2 and 13 in equation 20, one obtains for the phase velocity in the axial direction,

$$v_z=\frac{\tan\psi}{\sqrt{\mu_0\epsilon_2}}\sqrt{\frac{F_1}{F_2}} \tag{29}$$

where F_1 and F_2 are given by either equations 22 and 23 or by equations 24 and 25. The latter pair are for close wound helices. On the other hand if R_1/τ is sufficiently small, using equations 5 and 18,

$$v_z=\frac{1}{\sqrt{\mu_0\epsilon_2}}\sqrt{\frac{G_1}{G_2}} \tag{30}$$

where G_1 and G_2 are given by equations 27 and 28. In the event $\epsilon_1=\epsilon_2$ equation 30 reduces to the velocity of light in a medium characterized by the permeability μ_0 and the permittivity ϵ_2.

Some Examples

As an example consider a helical transmission line for which the dimensions are tabulated as follows:

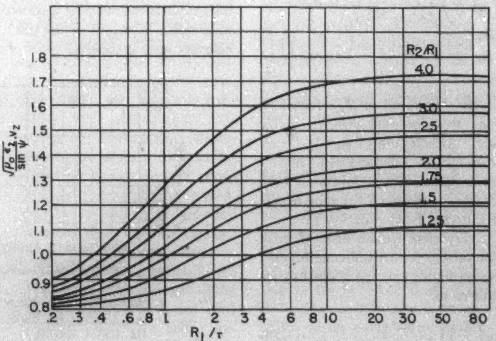


Fig. 6. Phase velocity versus R_1, τ, and R_2

Vol.
Pt. 1
195

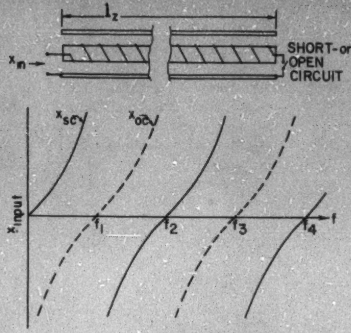


Fig. 7 (above). Test setup with typical results

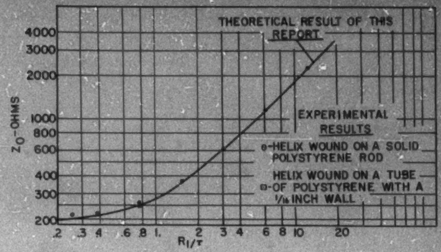


Fig. 8 (top right). Comparison of theoretical and experimental characteristic impedances

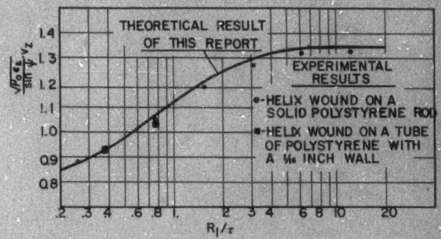


Fig. 9 (right). Comparison of theoretical and experimental phase velocities

$R_1=0.385$ inch
$R_2=0.750$ inch
$r=0.010$ inch
$\tau=$variable
$\epsilon_1=2.6\ \epsilon_0$
$\epsilon_2=1.0\ \epsilon_0$
$\epsilon_0=8.854\times10^{-12}fd/m$

These dimensions were chosen since they correspond to those for a series of helices tested by the author.

Fig. 4 shows the impedance as calculated from the close wound helix approximation which is embodied in equations 21, 24, and 25. This approximation was applied over the range of R_1/τ from 0.02 to 10.0. Of course, at the lower end of the range the helix can hardly be said to be close wound. Also in Fig. 4 is plotted the exact impedance as embodied in equations 21, 22, and 23. It is interesting to note that at $R_1/\tau=0.02$ they differ by only 8%.

Extensive families of curves have been calculated showing the characteristic impedance and phase velocity of wire helices. Typical examples of such families are shown in Figs. 5 and 6. For a more complete set of these curves, see reference 1.

Experimental Results

Other authors have conducted experimental work on the impedance and phase velocity of shielded helices. However, this author has conducted similar experimental work because the reports of the previous investigators left some doubts concerning the manner of testing or the actual dimensions, etc., of their helices. As will be shown, the work of this investigator agrees closely with the analytic results presented in the previous sections.

The method of testing used by the present author does not measure the characteristic impedance and phase velocity directly. These are determined from a measurement of the open-circuit and short-circuit driving point impedances of the transmission line. Fig. 7 shows the results which would be obtained from a typical set of measurements on a given lossless helical transmission line of axial length l_z. From transmission line theory applied to a lossless transmission line, the open- and short-circuit impedances are pure reactance. They are

$$X_{sc}=j\sqrt{\frac{L_z}{C_z}}\tan\left(2\pi f l_z\sqrt{L_zC_z}\right) \qquad (31)$$

$$X_{oc}=-j\sqrt{\frac{L_z}{C_z}}\cot\left(2\pi f_z\sqrt{L_zC_z}\right) \qquad (32)$$

where X_{sc} and X_{oc} are the short-circuit and open-circuit driving point impedances respectively. The parameters L_z and C_z are the distributed inductance and capacitance per unit of length in the axial direction. From equations 31 and 32,

$$Z_0=\sqrt{\frac{L_z}{C_z}}=\sqrt{|X_{sc}|\cdot|X_{oc}|} \qquad (33)$$

$$v_z=\frac{1}{\sqrt{L_zC_z}}=4f_1l_z \qquad (34)$$

where f_1 is the frequency of the first "pole" in the short-circuit driving point impedance. This frequency is also that at which the transmission line is one quarter of a wavelength long. Thus by measuring the open- and short-circuit driving point impedances for frequencies up to that at which the line becomes one quarter of a wavelength long, the characteristic impedance and phase velocity can be determined from equations 33 and 34.

The analytic results of the previous section were arrived at by assuming infinite wave length. Because of this, the experimental results are confined to frequencies for which the transmission line length is equal to or less than one quarter wavelength.

All of the helices tested had the same R_1, R_2 and wire radius r. These were respectively 0.385 inch, 0.750 inch, and 0.010 inch. The winding pitch was varied from 1/32 inch up to $1^1/_2$ inches. The outer conductor or sheath consisted of a 1.5-inch copper tube 72 inches long with a wall thickness of 1/8 inch. The first two helices, nos. 1 and 2, were wound on a polystyrene tube having a 3/4-inch outside diameter and a 1/16-inch wall. These helices were 72 inches long and were supported concentrically in the copper tube by three polystyrene supports placed at both ends and the middle. These supports were 1/8 inch thick and their effect was assumed to be negligible. The helix was made very long in order to

Table I. Comparison of Test and Analytical Results as Obtained by the Author

Helix Number	Pitch, Inches	$\frac{R_1}{\tau}$	Characteristic Impedance, Ohms		$\frac{\sqrt{\mu_0\epsilon_1}\cdot v_z}{\sin \psi}$	
			Calculated	Measured	Calculated	Measured
1	0.50	0.770	247*	248	1.068*	1.038
2	1.00	0.385	210*	216.2	0.930*	0.930
3	0.03125	12.32	2260	2250	1.347	1.328
4	0.0625	6.16	1140	1137	1.342	1.321
5	0.125	3.08	603	605.5	1.302	1.272
6	0.250	1.54	356	364.5	1.206	1.181
7	0.500	0.77	247	257.3	1.068	1.052
8	1.50	0.257	201	212.3	0.874	0.881

* Calculated for a solid core.

eliminate from consideration any end effects, the analysis being based upon an infinitely long helix. Helices 3 to 8 were the same as 1 and 2 in all respects except that they were wound on 3/4-inch solid polystyrene rod and were 71 inches long. Table I gives the $R_1\tau$ of the various helices along with the calculated and measured properties.

These results are plotted in Figs. 8 and 9. At the lower end of the range of R_1/τ, the measured impedances lie somewhat above the calculated impedances. This is probably because of the fact that the theoretical result is based upon the assumption that the dielectric material of the supporting rod extends out to a radius R_1. In actuality, it falls short of this by an amount equal to the wire radius. In Fig. 9 the phase velocity compares very favorably with the theoretical curve, the maximum deviation being 3%. One thing apparent from the test results is that it makes very little difference whether the helix is wound on a solid rod or upon a tube.

Conclusion

This paper has presented the results of a theoretical and experimental study of the properties of a shielded helical transmission line. It has been shown that the parameters of a shielded wire helix can be satisfactorily computed by replacing the actual equipotential surface of the wire by an equivalent surface. These results should prove useful in a number of applications where such a structure appears.

References

1. THE CHARACTERISTIC IMPEDANCE AND PHASE VELOCITY OF A SHIELDED HELICAL TRANSMISSION LINE, H. S. Kirschbaum. *Ph.D. Dissertation*, Carnegie Institute of Technology, Pittsburgh, Pa., 1953.

2. STATIC AND DYNAMIC ELECTRICITY (book). W. R. Smythe. McGraw-Hill Book Company, Inc., New York, N. Y., 1939.

3. A NOTE ON THE INDUCTANCE OF SCREENED SINGLE LAYER SOLENOIDS. F. M. Phillips. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 96, pt. 3, 1949, p. 138.

4. REFERENCE DATA FOR RADIO ENGINEERS, H. P. Westman, editor. International Telephone and Telegraph Corporation, New York, N. Y., fourth edition, 1956, p. 593.

Vol.
Pt. 1
1959

Precision Integrator for D-C Potentials

JOHN R. PATTEE
MEMBER AIEE

INTEGRATORS of small direct currents and potentials have been put to good use in a wide variety of applications, such as in measuring photographic exposure by time-integrating the output of a photovoltaic cell, in registering watt-hours derived from the d-c millivolt (mv) output of thermal watt converters and electronic watt converters, and in tallying the amount of gold consumed in gold-plating operations by integrating the potential drop in the plating current shunt.

General Description

A d-c integrator will be described which utilizes a d-c amplifier and a motor-driven tachometer d-c generator providing feedback potential as principal operating elements. The input circuit is free of make-and-break contacts and the amplifier gain is very high, so that a calibration accuracy of 0.1% is regularly achieved and potentials as low as 5 μv (microvolts) are resolved. Potentials may be of either polarity or may fluctuate back and forth through zero. Values of the integral are almost continuously registered by a mechanical counter directly driven by the motor generator, and the values increase or decrease according to the polarity of the input.

In this class of integrator, the basic process consists of amplifying the difference between the input potential to be time integrated and the output potential of a small d-c generator having a permanent-magnet field. The amplified difference controls the operation of a servomotor driving the generator. A fixed quantity of mv-seconds is generated at every revolution, being proportional only to the amount of flux cut by the revolving armature windings and the number of their turns, but independent of speed. When the generated mv-seconds are fed back in series opposition to the input mv-seconds, the two will sooner or later come into balance, because any remaining difference appears as voltage to drive the motor generator. Because servomechanisms have inertia and a dead zone, it is necessary to provide a surge capacitor and series resistor across the output of the d-c amplifier. The resistor is a mutual resistance between output and input circuits so that electricity is fed into the capacitor when the input potential is higher than the generator feedback, and is drawn out when the potential is lower. This action will be described in detail later in the paper.

Amplifier

It is generally known that the amplification of d-c potentials of a few mv with long-time stability presents serious problems because of thermal drift and noise. R. W. Gilbert has described a highly-stable d-c amplifier utilizing an induction galvanometer to convert small input currents to alternating current of high frequency.[1] This causes frequency modulation of an oscillator, the output of which is converted to bidirectional direct current in a discriminator. The input end consists simply of a galvanometer coil in series with the primary of a small 200-

---

Paper 59-166, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 5, 1958; made available for printing December 17, 1958.

JOHN R. PATTEE is with Central Research Laboratory, Weston Electrical Instrument Corporation, Newark, N. J.

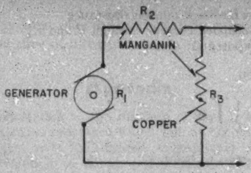


Fig. 1 (left). Temperature - compensated generator

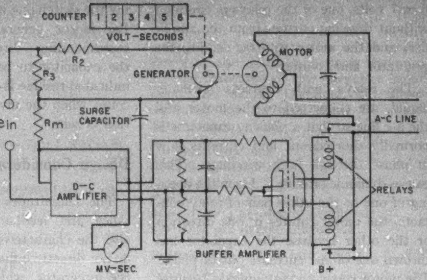


Fig. 2 (right). Schematic diagram

kc transformer. Over-all gain is very high, so that potentials as low as 5 μv are resolved.

This amplifier, named Inductronic,* is primarily responsible for the high accuracy and resolution achieved by the integrators described here. Accuracy is not materially affected by amplifier gain (as long as it is very high), nor by the magnitude of the mutual resistor or surge capacitor.

Generator

Accuracy also depends on the performance of the generator. This is a high-quality tachometer generator of a type which has been manufactured for many years so that its reliability in measurement service has been well established. The field magnet is of cobalt steel and brushes and commutator are of silver. It is operated at a speed far below that of normal service. A steel enclosure shields it from stray fields. Temperature compensation is provided to offset the negative temperature coefficient of the magnet, which is approximately −300 parts per million. Two methods have been applied. One consists of placing a magnetic shunt next to the magnet, the shunt being made of nickel-iron alloy with a negative coefficient of permeability. A similar method is used to compensate temperature influence on watt-hour-meter magnets. A difficulty in applying this method is that the shape and size of the shunt cannot be calculated, but must be determined by cut-and-try.

The second method consists of connecting a voltage divider across the generator output. As shown in the diagram of Fig. 1, the shunt branch has a section of copper wire. The positive coefficient of this section raises the proportion of the generated voltage appearing across that branch with rising temperature, while the field strength and generated voltage fall. The generator must be driven at sufficient speed to provide the voltage drop across the series branch as well as the required output voltage.

A mathematical analysis of generator temperature compensation is presented in the Appendix. The calculated values of resistance are only approximate because values of temperature coefficients which have been assigned to the components are not exact. The temperature coefficient of manganin is usually not quite zero as assigned, and the coefficient of the magnet may not be known to a high degree of accuracy. Nonetheless, a divider made according to the formula does yield a very low temperature coefficient of output. A more exact formula is not considered worthwhile, partly because there are other small temperature influences of unknown magnitude, such as dimensional changes. If high-precision compensation is required, a compensated generator is made according to the formula and tested over the required temperature range. If temperature influence is too high, it is a simple matter to make a small adjustment in the copper-to-manganin ratio of the shunt to correct it.

It is, of course, necessary that the temperature of the magnet, the armature winding, and the copper section of the shunt remain at the same temperature. Practically this means that they must be placed close together, shielded from concentrated external temperature variations, and substantially free of self-heating. All these requirements are met in the present design. The whole generator is enclosed in a steel shield. It has a cast-aluminum case completely enclosing magnet, armature, and resistors. All three resistors are wound on a small spool which is bolted to a part of the aluminum case close to the magnet and armature. Since the output is only 3 ma (milliamperes) at 270 mv, or 800 μw (microwatts), and the speed about 45 rpm, self-heating is negligible.

Detailed Description

The basic components so far discussed can be combined in several ways to form an integrator. The integrator which is the particular subject of this paper was briefly described in 1955 by R. W. Gilbert and will now be discussed in detail.[2] A simplified circuit diagram is given in Fig. 2. The block marked "d-c amplifier" represents the amplifier with induction galvanometer converter described earlier. The input circuit of the integrator consists of the amplifier input in series with a feedback resistor, R_m, and the generator, R_3. Current is fed back through the surge capacitor to the feedback resistor. The effect of this will be described in detail later. The potential across the capacitor is allowed to rise to several volts, and it (plus the small potential drop across the feedback mutual resistor) is applied to a voltmeter and to the grids of a dual triode which is a buffer between the amplifier and a pair of relays. The relays control the operation of a servomotor of low inertia which drives the generator through a 30/1 reduction gear. The motor, generator, and integral counter are connected by a chain drive. This lends flexibility to the design, so that various revolution ratios and various types of counters can be adapted without major mechanical changes. Operation of this assembly is explained in the following.

The potential to be integrated, applied to the input, produces a current output which flows through the surge capacitor and feedback resistor, rising instantly to such a value that the drop across the resistor very nearly balances the applied potential. (If the gain of the amplifier were infinite the balance would be perfect, in the absence of stray thermal electromotive forces (emfs).) An almost perfect current generator thus delivers to the surge capacitor, in proportion to the applied potential, a current which is not appreciably diminished by the rising back emf. The output potential of the amplifier goes to the grids of a dual triode having a relay in each plate circuit. When it has risen to some

---

* Registered trade-mark.

6 or 7 volts, one of the relays is operated without drawing current from the amplifier, and the motor rotates, driving the generator and counter.

The relays, with contacts normally closed, are connected to the motor and the a-c line so that a phasing capacitor is normally shorted out and currents flow in phase through both windings. This causes dynamic braking. Upon the opening of one or the other pair of relay contacts, the motor rotates in one direction or the other because the capacitor has thrown one of the currents out of phase. The motor starts up very quickly because of its low inertia.

The rotating generator delivers to its shunt a potential which, by design, is higher than the maximum rated steady-input potential and opposed to it. Then the net potential input at the amplifier is reversed, the output current reverses, and the surge capacitor is gradually discharged. At 2 or 3 volts, the contacts are closed and the motor stops suddenly with dynamic braking. In other words, current surges in and out of the capacitor, building up a potential which is an analog of the input potential-time integral, then drawing it down while the integral quantity is transferred to the counter mechanically. The cycling rate depends upon the input level, and is likely to be about 3 or 4 cps (cycles per second).

The mv-seconds produced by the generator must eventually match the mv-seconds of the source (except for a small residue represented by the charge in the surge capacitor), because any difference between them causes a current to charge the capacitor to the point where the motor generator starts up. It will run until the difference between the input mv-seconds and the quantity registered by the counter is very small. Thus, the integrating action of the counter is intermittent, but it does not lag the true value by more than 15 mv-seconds and the lag is likely to be less. When the integral has run up to 15 volt-seconds, the lag will be less than 0.1%.

When integral steps smaller than 15 mv-seconds must be resolved, a voltmeter is added which measures the analog voltage of the surge capacitor plus the small potential drop across R_m. This is the instrument, labeled "mv-sec" in Fig. 2, which indicates the residual integral not transferred to the counter. At the end of a run, the indicated value may be added arithmetically to the quantity registered by the counter, or it may be transferred to it electromechanically by operating a three-position switch, not shown, having contacts in series with the relays. Pushing the switch one way or the other causes the motor generator to rotate in a corresponding direction. This causes the capacitor to be discharged, and the indicated residue is correctly added to the count when the mv-second meter indication is zero.

Design Considerations

The magnitudes of the surge capacitance and feedback mutual resistance and the characteristics of the voltmeter do not directly influence the calibration of the integrator, which is governed solely by the generator. They are chosen to suit the characteristics of the Inductronic amplifier and the motor generator.

The generator, while running, delivers about 80 mv across the shunt, R_3. The average input potential must be below this value, so a normal average input of 70 mv has been established. The transfer characteristic of the amplifier is linear up to 1.8 ma of output. R_3 is made 100 ohms, which produces a balance of the 70-mv input with 0.7 ma, leaving an ample margin for momentary overloads.

Maintenance of transfer linearity also requires that the output potential be limited to 10 volts or less. This has a bearing on the choice of characteristics of both the voltmeter and the surge capacitor. When all other conditions are fixed, the cycling rate of the system is inversely proportional to the capacitance, C. A high cycling frequency keeps the residual integral value small but causes strain and wear of the motor and the mechanism it drives. A capacitance of 18 μf (microfarads) has been found to be a good compromise. Buffer amplifier gain and relay characteristics have been arranged to operate the relays at about 7 volts and release them at about 2 volts. The particular values are not important.

Calibrating the Millivolt-Second Meter

In view of these considerations, the full-scale range of the voltmeter is made 10 volts. A 10–0–10-volt zero-center scale is provided for reversible inputs, and a 0–10-volt scale for unidirectional systems. Although the voltmeter current, being outside the feedback loop, does not affect calibration, it is designed to be only 50 μa (microamperes) full scale in order not to overburden the amplifier. Scale calibration is in mv-seconds, and the characteristic can be determined analytically as follows, letting

e_{in} = input potential
C = capacitance of surge capacitor
R_m = feedback mutual resistance
i = feedback current through C and R_m
E_c = potential across C

Then

$$E_c = \frac{1}{C}\int i\,dt \tag{1}$$

$$iR_m = e_{in} \text{ at balance} \tag{2}$$

$$\int e_{in}dt = E_c R_m C \tag{3}$$

If $E_c = 10$ volts, $R_m = 100$ ohms, and $C = 18$ μf, $\int e_{in}dt = 18$ mv-seconds, this is the full-scale value. In practice, a complete voltmeter is procured, scaled 0–18 or 18–0–18 mv-seconds as required. The effects of tolerances in specified capacitance and voltmeter calibration are adjusted out if necessary by readjusting R_m, which does not affect the calibration of the counter. Calibration is unnecessary if the integral residue is transferred to the counter as described earlier, because the voltmeter then serves merely as a null indicator.

Capacitor Leakage

Calibration is not influenced by the currents flowing through the voltmeter and the buffer amplifier input because they are outside the feedback loop. But leakage current in the surge capacitor is a possible source of error, which can be analyzed as in the following.

Calling the capacitor resistance R_c, the leakage current will equal E_c/R_c. The potential drop it causes in the feedback resistor R_m is given by:

$$\text{Error potential} = E_c\frac{R_m}{R_c} \tag{4}$$

If the average value of E_c is 5 volts, R_m is 100 ohms, and R_c is as low as 100 megohms, the error potential will be 5 μv, which is about the resolving power of the system. Evidently capacitors can readily be obtained which would reduce this error to a negligible value.

Although the rated average input level is 70 mv, pulses of short duration amounting to twice that level are tolerated, provided that the time integral of the overvoltage does not exceed 200 mv-seconds.

Conclusions

An integrator is described for time-integrating direct potentials up to 70 mv average value of either polarity. Currents can be integrated by means of shunts, and power by means of watt converters. Potential resolving power is 5 μv. Integral values are registered with an accuracy of 0.1% or 1 mv-second, whichever is greater, plus the product of 5 μv times the time of integration.

Appendix

Referring to the diagram of Fig. 1, the values of resistance which will, at the same time, give the desired volt-seconds per revolution at the output terminals and a null temperature coefficient of output may be derived as follows, assuming that all the significant temperature coefficients are known:

M = temperature coefficient of magnet
C = temperature coefficient of copper
S = temperature coefficient of mixed shunt
t = a reference temperature of normal room value
Δt = a change in temperature from t
G = generated volt-seconds per revolution at temperature t
Q = output volt-seconds per revolution

Assume that manganin has no temperature coefficient.

At reference temperature t:

$$\frac{Q}{G}=\frac{R_3}{R_1+R_2+R_3} \tag{5}$$

If the temperature rises:

$$\frac{Q}{G(1+M\Delta t)}=\frac{R_3(1+S\Delta t)}{R_1(1+C\Delta t)+R_2+R_3(1+S\Delta t)} \tag{6}$$

$$Q[R_1+R_2+R_3+(R_1C+R_3S)\Delta t]=GR_3[1+(M+S)\Delta t+MS(\Delta t)^2]$$

From equation 1

$$Q(R_1+R_2+R_3)=GR_3$$

therefore

$$Q(R_1C+R_3S)\Delta t=GR_3[(M+S)\Delta t+MS(\Delta t)^2] \tag{7}$$

$MS(\Delta t)^2$, since it is a very small term, can be neglected.

$$QR_1C+QR_3S=GR_3M+GR_3S \tag{8}$$

$$S=\frac{GR_3M-QR_1C}{(Q-G)R_3} \tag{9}$$

To produce a temperature coefficient of S in the shunt R_3, it is made partly of copper, partly of manganin. The proportions must be:

$$\text{Copper resistance}=R_3\frac{S}{C} \tag{10}$$

$$\text{Manganin resistance}=R_3-\text{copper resistance} \tag{11}$$

In the present integrator, the generator is wound with copper; resistance across the brushes is approximately 12.3 ohms. The field magnet is treated so that the generated output, G, is 360 mv-seconds per revolution. One hundred mv-seconds has been chosen as a convenient terminal output, Q. It is desirable to limit the generator current in the divider network to a few milliamperes. But it is also desirable to keep the total internal resistance of the integrator input below 200 ohms. Therefore, the shunt resistance R_3 has been made 30 ohms. Tabulating the values:

G = 360 mv-seconds
Q = 100 mv-seconds
R_1 = 12.3 ohms
R_3 = 30 ohms
M = −0.0003 per degree centigrade
C = 0.004 per degree centigrade

Putting these values into equation 5, S = 0.00105/C.

From equations 10 and 11, the copper section should be 7.9 ohms and the manganin section 22.1 ohms.

The value of R_2 can be derived from equation 1:

$$R_2=R_3G/Q-R_1-R_3$$

$$R_2=65.7 \text{ ohms} \tag{12}$$

References

1. The Induction Galvanometer, a Sensitive Instrument Converter, R. W. Gilbert. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1121–26.

2. The Weston Model 1473 Precision Integrator, R. W. Gilbert. *Weston Engineering Notes*, Weston Electrical Instrument Corporation, Newark, N. J., vol. 10, no. 2, Dec., 1955.

Proposed Standard Test Codes for Magnetic Amplifiers

AIEE COMMITTEE REPORT

THE TEST CODES presented here were first presented in 1958 as a conference paper,[1] in substantially their present form. The slight revisions contained herein are the result of comments received since that presentation and as a result of a formal review of this paper by the Magnetic Amplifiers Committee.

It is intended that these proposed test codes be used as a guide for the industry in the obtaining and presentation of test data on magnetic amplifiers; naturally the number and type of tests selected will vary somewhat with the type and application of each particular amplifier.

As additional proposed test codes are formalized, they will be added to this list. The Test Codes Working Group is presently considering as subjects "Drift," and "Environmental Sensitivity," as well as test codes for bistable amplifiers.

Ratings

No work on magnetic amplifier ratings has been done since the presentation of a 1957 conference paper.[2] However, a separate working group is being formed this year to prepare standard ratings, including nominal ratings. For clarity and completeness in presenting the present proposed test codes, the following rating information has been extracted from the conference paper.[2]

> The rating of a magnetic amplifier consists of a statement of the values of the supply voltage, frequency, load resistance, output voltage or current, volt-ampere output, and control voltage or current assigned to it by the manufacturer. This rating shall be given for a specified connection of all components within the amplifier.
>
> . . . The magnetic amplifier shall be capable of operating continuously at [rated] supply voltage at rated frequency and with rated load . . ., for any values of output and control quantities within their ratings, without exceeding the established rating of any of its components.

Definitions

The magnetic-amplifier definitions contained in a proposed standard given in 1958[3] are adhered to. In addition, the following definitions are presented for clarity only; they are not a part of the proposed test codes, nor have they been formally approved by the Definition Subcommittee:

1. *Current amplification.* The current amplification of a magnetic amplifier is the ratio of the change in output current to the change in current in the control

---

Paper 59-177, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 5, 1958; made available for printing December 2, 1958.

The personnel of the Working Group on Test Codes Subcommittee of the Magnetic Amplifiers Committee are: H. L. Kellogg, *chairman*, H. W. Patton, F. G. Timmel (Working Group Chairman), H. B. Whitted, and T. G. Wilson.

The contributions of past, as well as present, members of the Standards Subcommittee of the AIEE Committee on Magnetic Amplifiers are gratefully acknowledged.

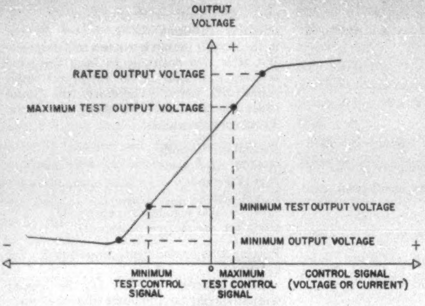


Fig. 1. Nonreversible parameters

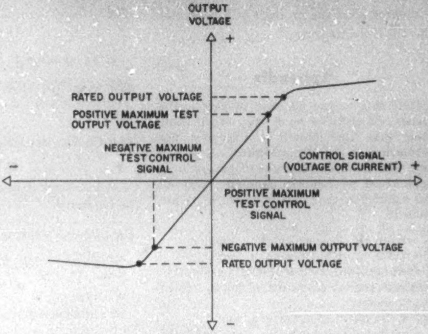


Fig. 2. Reversible parameters

winding required to produce the output current change.

2. *Figure of merit.* The figure of merit of a magnetic amplifier is the ratio of power amplification to the response time, preferably expressed in seconds.

3. *Power amplification.* The power amplification of a magnetic amplifier is the product of the voltage amplification and the current amplification.

4. *Voltage amplification.* The voltage amplification of a magnetic amplifier is the ratio of the change in output voltage to the change of voltage across the control winding circuit required to produce the output voltage change.

Proposed Standard Test Codes

I. GENERAL CONSIDERATIONS

A. *General Operating Conditions*

The magnetic amplifier should be operated at its nominal rated conditions except for its control circuit quantities and output quantities.

B. *Control Signal*

1. For the magnetic amplifiers with d-c signal, the control signal should be obtained from a voltage source which has less than 5% ripple and whose internal impedance is such that the source voltage at no time under transient or steady-state conditions decreases by more than 10% of the value which would exist if the control circuit of the magnetic amplifier were open.

a. The average magnitude of the control signal voltage should be measured at the terminals of the voltage source.*

b. Control circuit currents flowing as a result of this voltage should be measured with an average reading meter.*

2. For magnetic amplifiers with a-c signal, the control signal (carrier) should be obtained from a voltage source of power frequency and specified phase which has a total harmonic content less than 5% and whose internal impedance is such that the source voltage at no time under transient or steady-state conditions decreases by more than 10% of the value which would exist if the control circuit of the magnetic amplifier were open.

a. The magnitude of the average values (averaged over one-half cycle) of the a-c signal voltage should be measured.†

b. The average value (averaged over one-half cycle) of the control circuit currents flowing as a result of this voltage should be measured.†

3. Any impedance inserted between the external control circuit terminals of the magnetic amplifier and the terminals of the signal voltage source must be specifically described. This impedance, if present, must be the same in all of the tests described in the following.

C. *Output Voltage*

1. For magnetic amplifiers with d-c output, the average value of the output voltage should be measured.*

2. For magnetic amplifiers with a-c output, the output voltage should be measured in the same units as stated in the output rating of the amplifier.

a. For rms output ratings, the output voltage should be measured with a voltmeter which responds to the rms value.‡

b. For average output ratings, the output voltage should be measured with an instrument which responds to the average value.†

c. For ratings based on the fundamental component of the output voltage, the amplitude of the fundamental component should be measured with a wave analyzer.

D. *Output Current*

Output currents should be measured with the same type of instruments as are used to measure the output voltages.

E. *Load*

The amplifier shall be terminated in a resistive load or, if other loads are used, they must be accurately defined.

F. *The Determination of Parameters*

1. Rated output voltage is as specified in ratings.

2. Magnetic amplifier with nonreversible output:

a. Signal voltage should be applied to the control circuit of the magnetic amplifier to determine the minimum output voltage.

b. The maximum test output voltage is defined as the minimum output voltage plus 80% of the difference between the rated and minimum output voltages.

c. The minimum test output voltage is defined as the minimum output voltage plus 20% of the difference between the rated and minimum output voltages.

d. Minimum test and maximum test output currents are defined as those output currents which flow as a result of the corresponding output voltages.

e. Minimum test and maximum test control signal voltages are those control signal voltages required to give the corresponding output voltages.

f. Minimum test and maximum test control signal currents are defined as those control signal currents required to give the corresponding output voltages.

3. Magnetic amplifiers with reversible output.

---

* For example, a D'Arsonval meter movement can be used for steady-state d-c measurements.

† For example, steady-state a-c measurements can be made with instruments having full-wave rectification and D'Arsonval movements. Since these instruments are usually calibrated to measure the rms value of a pure sine wave, the scale reading should be divided by 1.11.

‡ For example, a thermocouple-type voltmeter can be used.

Vol.
Pt. 1
195

a. Positive maximum test output voltage is defined as 80% of the rated output voltage in the positive direction.

b. Negative maximum test output voltage is defined as 80% of the rated output voltage in the negative direction.

c. Positive and negative maximum test output currents are defined as those output currents which flow as a result of the corresponding output voltages.

d. Positive and negative maximum test control signal voltages are defined as those control signal voltages required to give the corresponding output voltages.

e. Positive and negative maximum test control signal currents are defined as those control signal currents required to give the corresponding output voltages.

4. Illustration of parameters:

a. Fig. 1 presents nonreversible parameters.

b. Fig. 2 presents reversible parameters.

II. Voltage Amplification

A. Magnetic Amplifiers with Nonreversible Output

1. The voltage amplification referred to any one control winding is the ratio of the difference between the maximum test output voltage and minimum test output voltage to the absolute value of the algebraic difference between maximum test control signal voltage and minimum test control signal voltage.

$$\frac{\text{max test output V}-\text{min test output V}}{|\text{max test CS V}-\text{min test CS V}|}$$

where V = voltage; CS = control signal.

2. Voltage amplification, as determined by the foregoing, is applicable not only when the control signal and output voltages are both a-c or d-c, but also when one is a-c and the other d-c.

B. Magnetic Amplifiers with Reversible Output

1. The voltage amplification referred to any one control winding is the ratio of the absolute value of the algebraic difference between positive maximum test output voltage and negative maximum test output voltage to the absolute value of the algebraic difference between the positive maximum test control signal voltage and the negative maximum test control signal voltage.

$$\frac{|\text{pm test output V}-\text{nm test output V}|}{|\text{pm test CS V}-\text{nm test CS V}|}$$

where pm = positive maximum; nm = negative maximum.

2. Voltage amplification, as determined by the foregoing, is applicable not only when control and output voltages are both a-c or d-c, but also when one is a-c and the other d-c.

C. Magnetic Amplifiers with Multivalued Control Functions

If, in determining the voltage amplification, it is found that the absolute value of the denominator of the ratio is less than the absolute value of the maximum excursion of the control signal voltage encountered in progressing from one test output voltage to the other, the absolute value of the maximum excursion of control signal voltage shall be used as the denominator.

III. Current Amplification

A. Magnetic Amplifiers with Nonreversible Output

1. The current amplification referred to any one control winding is the ratio of the difference between the maximum test output current and the minimum test output current to the absolute value of the algebraic difference between maximum test control signal current and the minimum test control signal current.

$$\frac{\text{max test o-p}-\text{min test o-p}}{|\text{max test CS current}-\text{min test CS current}|}$$

where o-p = output current.

2. Current amplification as determined by the foregoing, is applicable not only when the control and output currents are both a-c or d-c, but also when one is a-c and the other d-c.

B. Magnetic Amplifiers with Reversible Output

1. The current amplification referred to any one control winding is the ratio of the absolute value of the algebraic difference between the positive maximum test output current and the negative maximum test output current to the absolute value of the algebraic difference between the positive maximum test control signal current and the negative maximum test control signal current.

$$\frac{|\text{pm test o-p}-\text{nm test o-p}|}{|\text{pm test CS current}-\text{nm test CS current}|}$$

2. Current amplification as determined by the foregoing is applicable not only when the control and output currents are both a-c or d-c, but also when one is a-c and the other is d-c.

C. Magnetic Amplifiers with Multivalued Control Functions

If, in determining the current amplification, it is found that the absolute value of the denominator of the ratio is less than the absolute value of the maximum excursion of the control signal current encountered in progressing from one test output current to the other, the absolute value of the maximum excursion of the control signal current shall be used as the denominator.

IV. Power Amplification

The power amplification is the product of the voltage and current amplification when measured under identical operating conditions in accordance with sections II and III.

V. Response Time

A. Control Signals

The magnitude of the control signal must be the difference between the maximum test control signal and the minimum test control signal in the case of a magnetic amplifier with nonreversible output, or the absolute value of the algebraic difference between the positive maximum test control signal and the negative maximum test control signal in the case of a magnetic amplifier with reversible output.

B. Measurements

1. Determine the time (preferably in seconds; it may also be measured in cycles of the power frequency) required after application of the step-function signal for the output quantity to change by an amount equal to 63% (unless otherwise specified) of the final change in output.§ The response time shall be specified as the maximum obtained including such effects as growth or decay, and time phase of signal application.

2. The response time may also be obtained graphically or analytically from the frequency response; see section VI.

---

§ Because of the distorted output waveform obtained from magnetic amplifiers, special means are required to measure the time variation of the output quantity.

A. For magnetic amplifiers with rectified output, the time variation in the average output quantity may be measured in one of the following methods: 1. By using a cyclic integrator to measure the average value of each output half-cycle.[4] 2. By measuring the variation in the firing angle, with the use of oscillographic methods, and determining the variation in the output quantity by conversion factors derived from either steady-state measurements or analytical expressions.[5] 3. By using a suitable low-pass filter to suppress the supply frequency and its harmonics prior to oscillographic measurement. This method is applicable to magnetic amplifiers with relatively slow response (several cycles of the supply frequency).

B. For magnetic amplifiers with a-c output, the output quantity should be rectified and the average response determined as for a d-c output circuit. The average response can then be converted to either rms response or fundamental response by steady-state conversion factors.

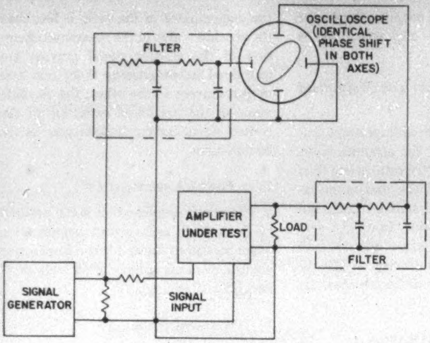


Fig. 3. Frequency response test circuit

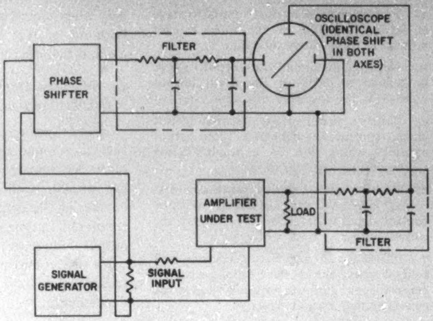


Fig. 4. Frequency response test circuit with phase-shifter

VI. Frequency Response: Amplification and Phase Versus Frequency

A. *Control Signals*

1. The control signals may be in the form of a low-frequency alternating current or in the form of modulation on a carrier. The modulating signal frequency should not exceed 25% of the amplifier's power supply frequency. This practical limit will ensure that the errors inherent in this method of measurement do not become too great.

2. The control signal amplitude should be held constant as the modulating frequency is varied. The control signal amplitude should be specified.

3. If the response time is to be determined graphically or analytically from the frequency response, the magnitude of the control signal must be the difference between the maximum test control signal and the minimum test control signal (measured with the signal frequency equal to zero) in the case of a magnetic amplifier with nonreversible output, or the absolute value of the algebraic difference between the positive maximum test control signal and the negative maximum test control signal (measured with the signal frequency equal to zero) in the case of a magnetic amplifier with reversible output.

B. *Measurements*

Determine the voltage or current amplification and the phase shift through the amplifier as a function of the modulating frequency.[‖]

C. *Presentation of Data*

Data may be presented by one or more of the following methods or their equivalents:

1. Graphing amplifier amplification and phase response versus modulating signal frequency.

2. Expressing the amplifier's performance using LaPlace or similar transforms.[¶]

VII. Figure of Merit

The figure of merit is the ratio of power amplification to the response time (preferably expressed in seconds) when measured under identical operating conditions in accordance with sections IV and V.

[‖] Magnetic amplifier output waveforms are distorted, and special means are required for proper measurements under dynamic conditions. Typical methods are shown in Figs. 3 and 4:

A. The d-c output from the amplifier (demodulation required for a-c carrier output) should be passed through a low-pass filter connected to the input terminals of an oscilloscope. The filter should suppress the carrier frequency and harmonics with as little disturbance to the signal frequency as possible.

B. The control signal should be applied through an identical filter to the other input of the oscilloscope. Demodulators will be required for carrier-type control signals. If a nulling method is used, as in Fig. 4, a phase shifter is required at the control signal source.

[¶] A transfer function of the following type may be obtained by "curve fitting" from the graphs referred to in paragraph 1 of section VI C. Such a linear transfer function is a very close approximation at signal frequencies which are well below the amplifier's power supply frequency:

$$G(s) = \left[\frac{K_1 e^{-st_1}}{1+T_1 s}\right]\left[\frac{K_2 e^{-st_2}}{1+T_2 s}\right]\cdots\left[\frac{K_n e^{-st_n}}{1+T_n s}\right]$$

where n = number of stages of amplification, t = dead time, and T = time constant. For a more rigorous mathematical treatment of transfer functions as applied to a pulse output device such as a magnetic amplifier, see references 6, 7, and 8.

References

1. Progress Report of the Standards Subcommittee of the AIEE Magnetic Amplifiers Committee, **AIEE Committee Report.** *AIEE CP 58-899* (available on request).

2. Progress Report of the Standards Subcommittee of the AIEE Magnetic Amplifiers Committee, **AIEE Committee Report.** *AIEE CP 57-98* (available on request).

3. Proposed Standard Terms and Definitions for Magnetic Amplifiers, **AIEE Committee Report.** *AIEE Transactions*, vol. 77, pt. I (*Communication and Electronics*), Sept. 1958, pp. 429–32.

4. The Cyclic Integrator—A Device for Measuring the Frequency Response of Magnetic Amplifiers, **T. Dunnegan, Jr., J. D. Harnden, Jr.** *Ibid.*, vol. 73, Sept. 1954, pp. 358–65.

5. Oscillographic Techniques for the Evaluation of Magnetic Amplifier Response, **D. L. Critchlow.** *Ibid.*, vol. 74, Nov. 1955, pp. 607–10.

6. Analysis of Magnetic Amplifiers by the Use of Difference Equations, **P. R. Johannessen.** *Ibid.*, vol. 73, 1954 (Jan. 1955 section), pp. 700–16.

7. Magnetic Amplifier Analysis and Application Using Block-Diagram Techniques, **Robert M. Hubbard.** *Ibid.*, vol. 76, Nov. 1957, pp. 578–88.

8. A Fast-Response Full-Wave Magnetic Amplifier, **G. E. Lynn.** *Ibid.*, vol. 77, Mar. 1958, pp. 37–41.

9. Progress Report of the AIEE Magnetic Amplifier Subcommittee, **AIEE Committee Report.** *Ibid.*, vol. 70, 1951, pp. 445–50.

Vol. 7
Pt. 1
1959

Long Time Delays From a Single Magnetic Storage Core

C. E. HARDIES
ASSOCIATE MEMBER AIEE

ACCURATE time delay devices are essential to both industrial and military control applications. Static time delay devices presently available provide time delays accurately and economically up to a few seconds. The emphasis on reliability in the industrial and military control fields has made it desirable that a static device be developed to generate the long time intervals required in these areas where reliability and accuracy are so fundamentally important.

This paper describes the theory and circuitry of a device which produces long time delay. It uses magnetic cores with essentially rectangular hysteresis loops, as idealized in Fig. 1, and transistors as switches. Accuracies of better than one per cent are achieved for time intervals less than 1/10 second up to several minutes. The device can be used in conjunction with static switching devices in performing the functions of logic for control purposes.

Some of the basic properties of the magnetic cores used in this device have been utilized in other circuit applications, and many effects other than amplification have been achieved. Dr. H. F. Storm[1] has written an excellent article summarizing some of the circuit applications. The property of a rectangular-loop magnetic core to absorb a fixed volt-second integral as determined by the flux change from one saturation level to the other has been used in short time delays by G. F. Pittman,[2] J. L. Lowrance,[3] R. L. Van Allen,[4] and others. The present device uses this basic integrating property in such a way that the time delays which can be achieved in a single stage are much longer than in previous devices. While time delays less than 1/10 second can be generated in the circuit, this device will have its main usefulness in the generation of long time intervals.

In this connection, another type of static timing device should be mentioned. These are based on resistance-capacitance (R-C) time constants. The use of R-C devices for short time delays offers an inexpensive, accurate device. To achieve accuracy for long time delays, the basic R-C device must be modified to such a degree that the advantage of its simplicity is lost.

In the device described by Pittman, this fixed volt-second value is used to advantage. By pulsing the magnetic core with unidirectional pulses of fixed volt-seconds, the core flux is changed from $-\Phi_s$ to $+\Phi_s$ in discrete steps as shown in Fig. 2. The time delay is determined by the rate and number of these volt-second pulses. This can be expressed as follows: where n is the number of pulses, N is the number of turns, Φ_s is the saturation flux density in lines, e is the the instantaneous value of of the core voltage applied for the time interval Δt.

$$n=\frac{2N\Phi_s}{\int_{\Delta t} e dt}\times 10^{-8} \qquad (1)$$

The present device extends this basic reasoning to include the proposition that this voltage function need not be unidirectional. Time delays are achieved by alternately pulsing the core with positive and negative volt-second pulses. The difference between the positive and negative pulses is adjusted to be a few microvolt-seconds. The time interval which can be generated is determined by the number of times this difference must be applied to the core in order to change the flux level from one saturation to the other. The number of full cycles which must be applied to the core to accomplish this flux change can be expressed:

$$n=\frac{2N\Phi_s}{\text{(difference volt-seconds)}} \qquad (2)$$

The smaller the difference between the positive and negative pulses, the greater the values of n. Experimentally, it is not difficult to achieve values of n in excess of 7,000 with a single storage core. Since this number can be made so large, the pulses can be taken directly from a 60-cycle power source with time delays of 2 minutes using a single magnetic core. Using unidirectional pulses, four stages would be required to produce an equivalent time interval.

There are other advantages gained using the reversing polarity pulses:

1. The circuitry can be symmetrical, thus minimizing the effects of temperature, voltage, and frequency. The idealized hysteresis loop in Fig. 1 shows a constant magnetizing current, however, an actual hysteresis loop for a practical core will show a magnetizing current which varies as the core flux level is changed. Symmetrical operation will minimize the effects of this variation.

2. Another effect of magnetizing current which can affect the accuracy of these timing circuits is the volts-per-turn dependency of such currents. By symmetrical operation the error caused by volt-per-turn changes is almost eliminated.

3. It is possible to supply the device directly from the 60-cycle power source. This eliminates the need for special pulse forming circuits or low frequency generators.

Basic Circuitry

The basic theory of the device requires that the voltages applied to the core on alternate half cycles of the power source be unequal. This can be achieved in at least three ways: a current control, a volt-

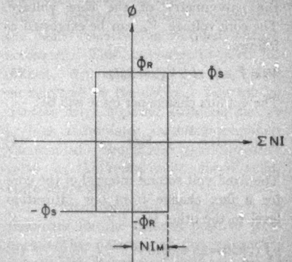


Fig. 1. Idealized rectangular hysteresis loop

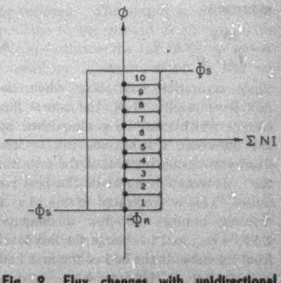


Fig. 2. Flux changes with unidirectional pulses

Paper 59-176, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 3, 1958; made available for printing December 4, 1958.

C. E. Hardies is with Magnetics, Inc., Butler, Pa.

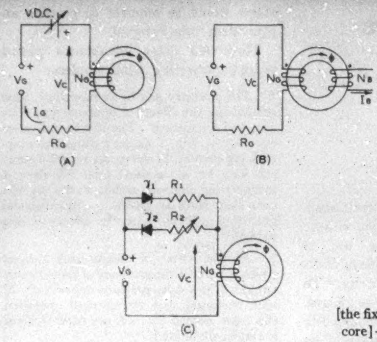


Fig. 3 (left). Basic timing circuits

A—D-c voltage control
B—D-c current control
C—Resistive control

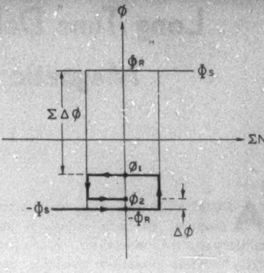


Fig. 4 (right). Flux changes with assymmetric bidirectional pulses

age control, or a resistive control. The effect of this control is the same in all three, making it possible for combinations to be used in the same application. Basic circuits for the three methods are shown in Fig. 3. Each of the three will be described briefly, and then a method by which delayed positive feedback can be used to provide the necessary output level for each.

Fig. 3(A) shows a circuit where a d-c voltage is used to produce and control the assymmetry of the core voltage. The core voltage, V_c, can be expressed as follows:

$$V_c = \hat{V}_g \sin \omega t + V_{\text{d-c}} = R_g I_g \tag{3}$$

The net flux change per cycle will be:

$$\phi \text{ per cycle} = \frac{10^8}{N} \int_{\Delta t} V_c dt \tag{4}$$

The fixed volt-second intergal of the core for a flux change from one saturation level to the other is:

$$n \int_{\Delta t} V_c dt = 2N\Phi_s \times 10^{-8} \tag{5}$$

The number of cycles of operation of the circuit of Fig. 3(A) which will cause a flux change from $-\Phi_s$ to $+\Phi_s$ can be expressed:

$$n = \frac{2N\Phi_s}{\int_{\Delta t} V_c dt} \times 10^{-8} \tag{6}$$

Since saturation will occur when the flux level reaches $+\Phi_s$, the actual flux change which must be accomplished by the difference volt-seconds will be the fixed volt-second integral of the core less the volt-second integral of the positive pulse. This is illustrated in Fig. 4. A discreet number of flux increments, $\Sigma\Delta\Phi$, is required to change the flux level from its value at the end of the first half cycle to $+\Phi_s$. This can be expressed as follows:

$$n = \frac{\text{[the fixed volt-second integral of the core]} - \text{[volt-second integral of the positive half cycle of the core voltage]}}{\text{[full cycle net volt-second integral of the core voltage]}} \tag{7}$$

$$n = \frac{2N\Phi_s \times 10^{-8} - \int_0^{\pi} (\hat{V}_g \sin \omega t + V_{\text{d-c}} - R_g I_g) dt}{\int_0^{2\pi} (\hat{V}_g \sin \omega t + V_{\text{d-c}} - R_g I_g) dt} \tag{8}$$

The value of n can be found readily if I_g is known as a function of time. This value of I_g will be dictated by the hysteresis loop assumption. If the voltage drop $R_g I_g$ is several orders of magnitude less than V_g, then the magnitude of I_g can be assumed equal to I_m of the assumed loop over the entire half cycle. With loops as displayed by Orthonol (50–50 nickel-iron alloy) cores, the symmetrical operation of the circuit will permit the use of an average value for I_m and thus, a constant value of I_g over the entire half cycle.

Integration of equation 8 yields:

$$n = \frac{2N\Phi_s \times 10^{-8} - \frac{2\hat{V}_g}{\omega} + (V_{\text{d-c}} - R_g I_g) \frac{\Delta t}{2}}{V_{\text{d-c}}(\Delta t)} \tag{9}$$

This can be converted to time delay for a 60-cycle-per-second frequency.

$$T = \frac{2N\Phi_s \times 10^{-8} - \frac{2\hat{V}_g}{\omega} + (V_{\text{d-c}} - R_g I_g) \frac{\Delta t}{2}}{V_{\text{d-c}}} \tag{10}$$

where $\Delta t = 1/60$ and $\omega = 377$.

As seen by equation 10, the delay time is dependent on voltage and frequency. By using a core and number of turns which will make the fixed volt-second integral of the core 20 times the volt-second integral of the positive volt-second-pulse, accuracy of one per cent will be achieved for a combined voltage and frequency change of ±10% each. A 20 to 1 relationship between the fixed volt-second integral of the core and the volt-second integral of the positive half cycle is easily achieved in practice. Experimental results are given using this ratio.

Fig. 3(B) shows a circuit which uses a d-c bias on the core to produce the assymmetrical core voltage. In this circuit configuration the core is biased to a magnetomotive force by the applied bias current I_B. This will cause the primary current I_g to assume different values on positive and negative half cycles and thus, a different $R_g I_g$ voltage drop. The expression for n, the number of cycles of operation before saturation in this case, can be expressed by equation 7. Substitution into this equation yields:

$$n = \frac{2N\Phi_s \times 10^{-8} - \int_0^{\pi} (\hat{V}_g \sin \omega t - R_g I_g) dt}{\int_0^{\pi} (\hat{V}_g \sin \omega t - R_g I_g) dt + \int_{\pi}^{2\pi} (\hat{V}_g \sin \omega t - R_g I_g) dt} \tag{11}$$

The value of I_g can be expressed, as shown in Fig. 5, by

$$I_g]_{0\,\mathrm{TO}\,\pi} = I_m - I_B \left(\frac{N_B}{N_G} \right) \tag{12}$$

$$I_g]_{\pi\,\mathrm{TO}\,2\pi} = I_m + I_B \left(\frac{N_B}{N_G} \right) \tag{13}$$

where $I_g R_g$ is again several orders of magnitude less than $\hat{V}_g$.

Integration of equation 11 with the values of I_g from equations 12 and 13 yields:

$$n = \frac{2N\Phi_s \times 10^{-8} - \frac{2\hat{V}_g}{\omega} - R_g \left[I_m - I_B \left(\frac{N_B}{N_G} \right) \right] \frac{\Delta t}{2}}{2R_g I_B \left(\frac{N_B}{N_G} \right) \frac{\Delta t}{2}} \tag{14}$$

In terms of time this equation is:

$$T = \frac{2N\Phi_s \times 10^{-8} - \frac{2\hat{V}_g}{\omega} - R_g \left[I_m - I_B \left(\frac{N_B}{N_G} \right) \right] \frac{\Delta t}{2}}{R_g I_B \left(\frac{N_B}{N_G} \right)} \tag{15}$$

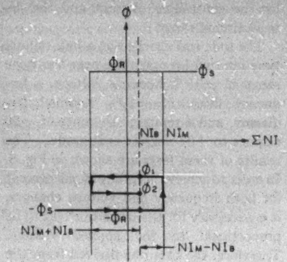


Fig. 5. Flux changes with d-c bias current and symmetrical pulses

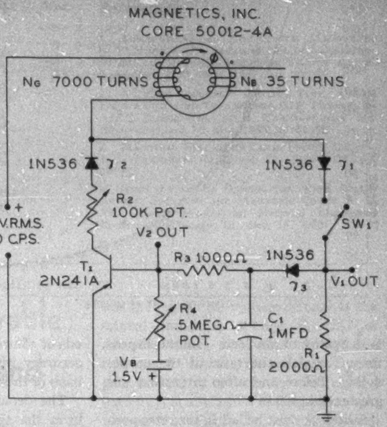


Fig. 6. Single-stage time delay

As in the case of the d-c voltage control, a 20 to 1 ratio must exist between the fixed volt-second integral of the core and the volt-second integral of the positive half cycle in order to maintain one percent accuracy for combined voltage and frequency variations of ±10% each.

In Fig. 3(C) a method of control is shown where the core voltage assymmetry is produced and controlled by resistive means. If the rectifiers γ1 and γ2 are assumed perfect, then the resistive voltage drop on the positive half cycles of the applied voltage will be R_1I_g, and on the negative half cycle R_2I_g. The net volt-second difference will be:

$$\int_0^{\pi} (\hat{V}_g \sin \omega t - R_1 I_g)dt + \int_{\pi}^{2\pi} (\hat{V}_g \sin \omega t - R_2 I_g)dt \quad (16)$$

Assymmetry is produced by the difference between the values of R_1 and R_2. The time in seconds is:

$$T = \frac{2N\Phi_s \times 10^{-8} - \frac{2\hat{V}_g}{\omega} - R_1 I_g\left(\frac{\Delta t}{2}\right)}{\frac{(R_2 - R_1)I_g}{2}} \quad (17)$$

Again, the 20 to 1 ratio between the fixed volt-second integral of the core and the volt-second integral of the positive half cycle must be maintained for one percent accuracy with a combined frequency and voltage variation of ±10% each.

Comparison between equations 10, 15, and 17 shows that all three will produce a similar characteristic, delay time versus direct voltage, direct current, or resistance. Any of the three methods can be used to establish the delay time.

In all of the cases just described, saturation is indicated at the end of the time interval by a sharp increase in the voltage drop across the series resistor. This voltage drop occurs near the end of the half cycle and does not represent much power. High output power is achieved by removing the negative half cycle once saturation is indicated. This will cause the core to self-saturate. A transistor circuit can be used for this purpose.

The circuit of Fig. 6 shows a circuit using resistive control, with a transistor for attaining output power. If resistors R_1 and R_2 are made equal, any of the control methods can be used. The transistor, T_1, of Fig. 6 is biased normally in its "on" or short circuit condition. This bias is somewhat less than the voltage across R_1 after saturation occurs. The diode γ_3 is provided to transfer the voltage across R_1 to the R-C network R_3 and C_1. The resistor R_3 is used for protection of the transistor's base. The capacitor C_1 provides a phase shift between the voltage R_1I_g and the voltage which appears across the base to emitter terminals of the transistor, T_1. This must be sufficient to hold the transistor open-circuited during some part of the succeeding half cycle. If this is achieved, the core will have a smaller volt-second reset during this half cycle and the output in the next positive half cycle across R_1 will increase. This "avalanche" action will continue until the transistor is held open circuit for the entire negative half cycle and the core is self-saturated. Very little voltage will be absorbed by the core, and the output will be half-wave rectified.

This avalanche or delayed positive feedback can take many cycles of the excitation before full output is achieved. The amount of time required can be used to compensate the device for inaccuracy due to temperature since the number of cycles during avalanche is a function of the bias on the transistor. An increase of temperatures on the core shortens time delays, while the same temperature change on the transistor lengthens them to a greater degree. These effects tend to self-compensate, but the bias supply to the transistor must still be temperature compensated to achieve the required accuracy.

The output of the device has been described as half-wave rectified across resistor R_1. Across the capacitor C_1, there exists a filtered d-c which is essentially the peak value of the rectified voltage across R_1. This may also be used for output. Approximately 1/2 watt can be delivered by the device. This can be used to drive a power switching device such as a transistor, magnetic amplifier, vacuum-tube amplifier, high-impedance relay, or the recently developed solid-state thyratron.

The circuit shown in Fig. 6 uses a p-n-p transistor for the delayed positive feedback. The polarity of the output is negative to the ground as shown in the figure. By reversing the diodes γ_1, γ_2, and γ_3 and using an n-p-n transistor for avalanche, the polarity of the output can be reversed. The output would then be positive to the ground as shown. This would facilitate the use of p-n-p power transistors or other output devices requiring this polarity signal.

The switch, S_1 in Fig. 6 is used to reset the core to its $-\Phi_s$ level after the timing interval. When switch S_1 is open, only those half cycles of the applied voltage, V_g are applied to the core which provide negative or resetting volt-seconds. When S_1 is open the output is zero. Closure of S_1 initiates the timing cycle. While this is taking place, the voltage across resistor R_1 and capacitor C_1 is very low. After saturation and avalanche, the voltage across these points rises, and output

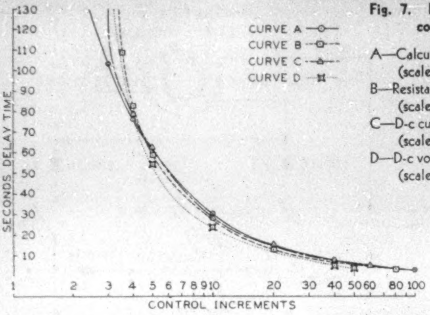


Fig. 7. Delay in second versus control increments

A—Calculated (scale × 0.01 = d-c volts)
B—Resistance controlled (scale × 100 = [$R_2 - R_1$] ohms)
C—D-c current controlled (scale × 1 = milliamperes)
D—D-c volts controlled (scale × 0.01 = d-c volts)

is delivered to the load. In the experimental models the ratio of the output voltage before and after saturation was greater than 100 to 1.

Switch S_1 must be left in the reset position for enough cycles to allow the core flux to be completely reset to its $-\Phi_s$ level before the device can be recycled. This will amount to 20 cycles if the ratio between core fixed volt-second capacity and applied half-cycle volt-seconds is maintained at 20 to 1 as previously determined.

The device could be used for purposes other than time delay. It could be a frequency divider if a transistor replaced switch, S_1, and a bistable device is used to drive this transistor. Signal for the bistable device can be taken from a resistor in series with the magnetic core. The voltage drop across this resistor is essentially zero during timing intervals and has the same polarity as the voltage across R_1 after the timing cycle. During reset it is again essentially zero and after the core has been reset, it has the opposite polarity as at the end of the timing cycle.

The high sensitivity of the device at long time intervals suggests its application in converting low-level analog signals to frequency. The core storage capacity could be designed in order to achieve high sensitivity for any desired frequency range.

Experimental Results

The operation of the timer circuit of Fig. 6 was experimentally checked for all three methods of establishing time intervals. These plots are shown together with the theoretical characteristic plot for the timer. As previously noted, this characteristic plot is common to all three methods of control. The scale for the independent variables was determined by use of equations 10, 15, and 17. The similarity of the plots shows the feasibility of combining the different methods of control in the same device and the accuracy which can be predicted on the basis of these equations.

The departure at long time intervals from the theoretical curve was caused by the transistor. When the curves were plotted without this transistor, and thus without sufficient output, the experimental curves followed the calculated one quite closely. The use of the transistor circuitry has permitted an output power level which can be utilized to drive switching devices. Its inclusion, however, has introduced difficulties in making a device which is continuously variable over the entire delay range while maintaining high accuracy.

Several sources of error were discovered while the experimental plots of Fig. 7 were being made. The error in reproducing readings of a given interval was approximately 0.5%. Several readings were taken of each time interval, and averaged to minimize this 0.5% error.

The d-c current method of setting time intervals was found to deviate considerably from the predicted curve for large control currents. Actually the current bias could only be plotted to 25 milliamperes. It was found that the bias ampere turns became equal to the magnetizing current ampere-turns at this level, and the core flux could be changed due to the bias ampere-turns alone.

The plots of Fig. 7 indicate that a given core volt-second capacity will provide a given range of delay times. The range for the magnetic core used in the experimental model extends from approximately 5 seconds to 100 seconds since, for times in excess of 100 seconds, the adjustment of the control becomes quite critical. In order to make the device less sensitive to control changes, the core volt-second capacity could be increased or two cores placed in series. The choice of core volt-second capacity must be made in view of both the sensitivity required by the controlling element and the desired timing range.

The unit was checked at a half minute time interval for operation over a voltage range of ±10% nominal voltage, a frequency range of ±10% nominal frequency, and a temperature range of −20 degrees to +60 degrees centigrade. The results of these tests are shown in Fig. 8. In order to achieve accuracy of one percent for both frequency and voltage changes, it is necessary for the transistor bias to be proportional to the applied voltage. Therefore, the bias was derived from the a-c excitation to obtain the accuracies shown in Fig. 8. The device was subjected to both shock and vibration in an effort to determine the effect on the timing interval. Preliminary results show the effects of vibration and shock will be small. Quantitative results were not available.

Conclusions

Time delays up to several minutes can be accurately generated in a single stage utilizing solid state components only. Experimental models have been built which will provide specific time delays up to 2 minutes with accuracy better than one per cent.

Effects other than time delay can be realized using this circuitry; for example, counting, frequency division, analog signal to time, or analog signal to frequency.

The signal required for the device can be: a d-c voltage, a d-c current, a change of resistance, or a combination of these.

The number of counts which can be

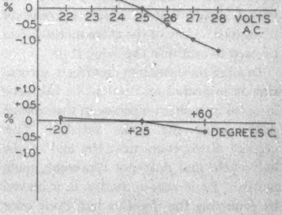


Fig. 8. Per-cent accuracy versus frequency, voltage, and temperature for a fixed 30-second delay

reliably stored in a rectangular hysteresis-loop magnetic core can be predicted using the derived equations.

Large numbers of counts can be stored in a single core by using bidirectional pulses. This use of bidirectional pulses minimizes the problems caused by core characteristics.

Nomenclature

n = number of cycles per time interval
N = turns
$\pm\Phi_s$ = magnetic flux at saturation in Maxwells
e = instantaneous voltage
$\Delta\phi$ = net core flux change per cycle
v_g = peak value of applied sinsuoidal voltage
$V_{d\text{-}c}$ = d-c control voltage
R_g = lumped gate circuit resistance
I_g = average gate current
Δt = period of one full cycle
N_G = gate windings
N_B = bias windings
I_B = average bias current
I_m = average magnetizing current
$V_{1\,\text{out}}$ = full-wave rectified output
$V_{2\,\text{out}}$ = filtered d-c output

References

1. Applications of Non-Linear Magnetics, H. F. Storm. "Proceedings of the Technical Conference on Magnetic Amplifiers," *AIEE Special Publication T-98*, Sept. 1957, pp. 205–29.

2. A High-Accuracy Static Time Delay Device Utilizing Transistors, G. F. Pittman, Jr. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, Mar. 1955, pp. 54–58.

3. Magnetic Saturable-Core Timing Device, J. L. Lowrance. *Ibid.*, vol. 77, July 1958, pp. 393–97.

4. Time Interval Generation with Transistors, Switches and Magnetic Cores, R. L. Van Allen. *Proceedings*, National Electronics Conference, Chicago, Ill., Oct. 15, 1958.

Analysis of Series-Connected Saturable Reactor with Capacitive Loading and Finite Control Resistance by Use of Difference Equations

H. C. BOURNE
MEMBER AIEE

J. T. SALIHI
ASSOCIATE MEMBER AIEE

Synopsis: A series-connected saturable reactor with capacitive loading exhibits interesting features. The circuit may exhibit instability associated with a double or three-valued output corresponding to a given input depending on previous history of operation. In some ranges of circuit parameters and forcing function, the circuit may become regenerative thus oscillating at a low frequency modulating the carrier.

A brief survey of the different phenomena occurring in the circuit is given in the first part of this paper. The rest of the paper is devoted to a thorough analysis of the oscillatory feature of the circuit. Assuming a rectangular forcing function, the critical conditions for oscillation and the frequencies at which the circuit oscillates are derived.

Since a saturable reactor essentially responds to the half-cycle average input, it may be expected that the theoretical results obtained for the case of a rectangular forcing function predict approximately the behavior of the circuit for the case of sinusoidal excitations. A comparison of measured results obtained from a circuit excited by a sinusoidal voltage with the theoretical results deduced in this paper show good correlation.

A REVIEW of the literature indicates that very little work has been done concerning the operation of magnetic amplifiers with capacitive loading. Addition of capacitance to magnetic amplifier circuits, modifies their modes of operation, and adds to the complexity of the analysis involved due to discontinuities in current or jump phenomena and low-frequency oscillations.

A practical investigation of common magnetic amplifier circuits with capacitive loading reveals interesting behavior, particularly the series-connected saturable reactor circuit shown in Fig. 1. This circuit consists of two identical reactors with primary and secondary windings, which are connected in such a way that a voltage applied to the two primary (gate) windings in series produces zero current in the secondary (control) winding as indicated by the polarity marks. The behavior of the circuit changes appreciably with the amount of resistance in the control circuit. Thus it is convenient to divide the operation into two cases of infinite and finite control impedance as follows.

Infinite Control Impedance. Assuming a sinusoidal gate voltage, the steady-state analysis of a series-connected saturable reactor with capacitive loading and infinite control impedance is given in reference 1. The behavior of the circuit is described by the nondimensional control characteristics of Figs. 2(A) and (B). The characteristics of Fig. 2(A), relating the nondimensional control current $R_g\bar{I}_c/\sqrt{2}E$ to the nondimensional gate current $R_gI_g/\sqrt{2}E$, are applicable when the nondimensional gate voltage $E/AB_cN_g\omega$ is less than a critical voltage defined as the critical transition voltage $E_t/AB_cN_g\omega$. At such values of applied voltage the relationship between the average value of gate current I_g and the referred control current I_c is a function of the phase angle ($\phi=\tan^{-1} 1/\omega CR_g$) only. Thus assuming a phase angle of 80 degrees, it may be seen from Fig. 2(A) that as the control current is increased from zero the average value of gate current increases linearly along the line OM. At M the gate current reaches its maximum value where, if the control current is further increased, the gate current drops abruptly to A_1 and remains at a constant value for further increase in control current. However, if the control current is reduced the gate current maintains its constant value only until the point C_1 is reached then it increases smoothly to the point D_1 where for further decrease in control current the gate current jumps to the point E_1. Then the relationship is again linear as described by the line OM, so that if the control current is increased the gate current increases linearly to M. Then it repeats the previous cycle. But if the control current is reduced, the gate current reduces linearly to zero. For $\phi=70$, 60, 50, and 40 degrees, the same sort of operation occurs. Thus for $\phi=70$ the operation is described by the path $OHA_2C_2D_2$-E_2O. The bistable operation disappears at a theoretical value of phase angle of 32.5 degrees. For phase angles less than this value the operation is reversible as seen from the curves corresponding to $\phi=0$, 10, 20, and 30 degrees, where the deviation from the linear relationship occurs before the point M is reached, that is, at the

---

Paper 59-170, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted August 6, 1958; made available for printing December 2, 1958.

H. C. Bourne is with the University of California, Berkeley, Calif., and J. T. Salihi, formerly with the Lenkurt Electric Company, San Carlos, Calif., is now in Iraq.

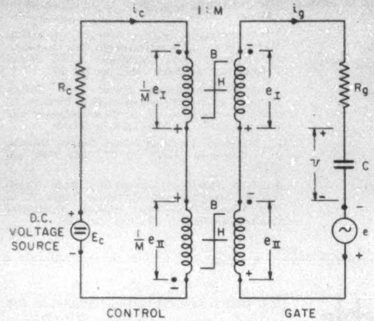


Fig. 1. Idealized series-connected saturable-reactor circuit with capacitive loading and finite control resistance having ideal reactors of zero-winding resistance and the magnetization characteristic of Fig. 5

points P_9, P_8, P_7, and P_6, respectively. For $\phi=0$, corresponding to the case of resistive loading discussed in reference 2, the control characteristic is described by the path OP_9QR.

When the value of the nondimensional applied voltage $E/AB_cN_g\omega$ exceeds the critical transition voltage, the control characteristics become a function of both the phase angle and the applied voltage. The control characteristics corresponding to a phase angle of 70 degrees are given in Fig. 2(B). The value of the critical transition voltage $E_t'/AB_cN_g\omega$ at this value of phase angle is given as 0.61. Thus for $E/AB_cN_g\omega=1.11$, which exceeds the critical transition voltage, examination of Fig. 2(B) shows that as the control current is increased from zero, the average value of gate current increases linearly from zero along the line OM up to the point B_3, where the current suddenly jumps to C_3. If the control current is then reduced the operation is along the path $C_3S_3A_3O$, while if the control current is increased the operation is along the path C_3D_3. At D_3 the gate current suddenly drops to the point E_3, where for further increase of control current the gate current remains constant as indicated by the horizontal line E_3L. However, if the control current is reduced the gate current remains constant only until the point F is reached, then it rises to the point G smoothly. When G is reached, the gate current rises abruptly to the point J_3, where, if the value of control current is reduced, the operation is described by the path J_3C_3-S_3A_3O; while if the control current is increased the path $J_3D_3E_3L$ is followed. The same sort of operation occurs for $E/AB_cN_g\omega=0.91$, as seen from the

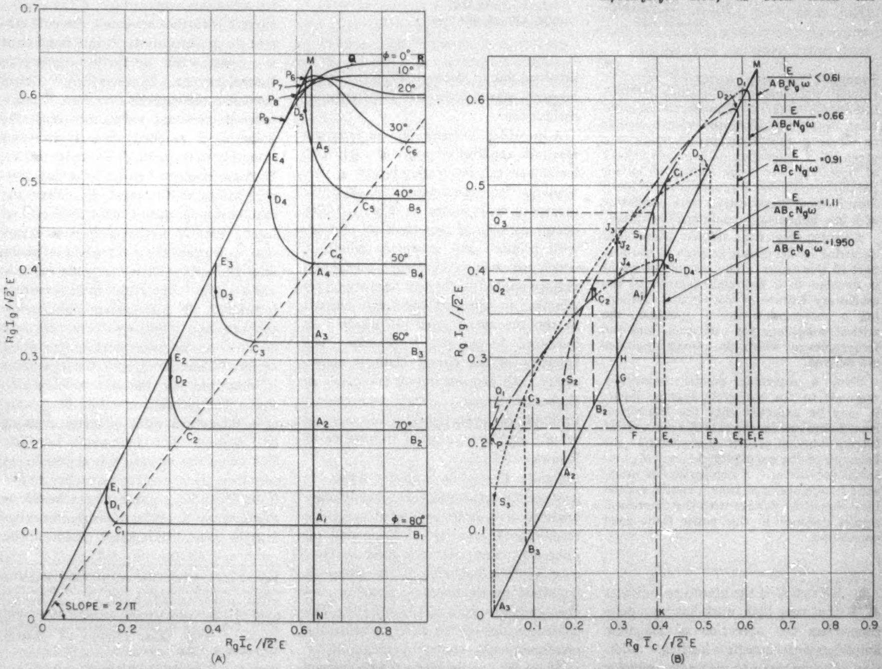


Fig. 2. Control characteristic of the series-connected saturable reactor of Fig. 1 supplied with a sinusoidal gate voltage and controlled by a d-c soucre, for values of applied voltage smaller (A), and larger (B), than the critical transition voltage defined in reference 1

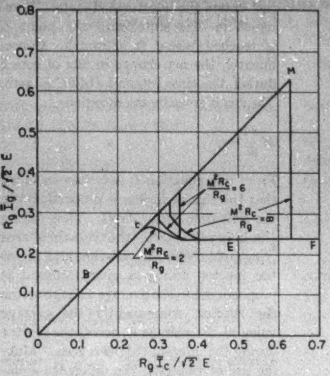


Fig. 3. Measured relationship between gate and control current taken at three fixed values of $M^2R_c/R_g=\infty$, 6, and 2, for the series-connected saturable reactor of Fig. 10 using two reactors of Fig. 4 supplied from a sinusoidal gate voltage and controlled by a voltage source. In these three cases ϕ was adjusted to a fixed value of 70 degrees. The case of infinite-control impedance was approximated by inserting a high inductance in series with the control windings. For $M^2R_c/R_g=6$ and 2, the corresponding values of R_c were 2,400 and 800 ohms, respectively, $E=40$ volts, $f=60$ cps

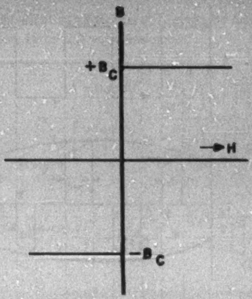


Fig. 5. Assumed ideal magnetization characteristic

corresponding curve of Fig. 2(B). However, for $E/AB_cN_g\omega=0.66$ the operation differs slightly from the previous cases, since there is not point corresponding to J_3 and J_2. Thus, when the point G is reached, the gate current shows only a small jump to the point H, where the relationship between the gate and control current becomes linear. Also for $E/AB_cN_g\omega=0.66$, the vertical line C_1B_1 is to the right of the point H, consequently for a given value of control current corresponding to the point K, the gate current might take one of the three possible values Q_1, Q_2, or Q_3 depending on the history of operation memorized by the circuit. For the other curves there are only two possible values of gate current for a given value of control current.

When values of $E/AB_cN_g\omega$ exceed a second critical voltage defined as the critical extinction voltage $E_c'/AB_cN_g\omega$ the behavior is different in that the gate current does not reduce to zero as the control current is reduced. The value of the critical extinction voltage is larger than the critical transition voltage, and is a function of the phase angle. For example the value of the critical extinction voltage $E_c'/AB_cN_g\omega$ at a phase angle of 70 degrees is about 1.14. Thus for $E/AB_cN_g\omega=1.950$, which exceeds the critical extinction voltage, the behavior is described by the path $PD_4E_4FGJ_4P$, shown in Fig. 2(B). Consequently, for magnetic amplifiers with capacitive loading, a condition of normal or under excitation may be defined corresponding to the case of resistive loading[2] by restricting the magnitude of the applied voltage to a value less than the critical extinction voltage.

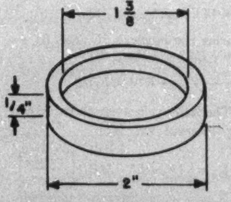


Fig. 4. Core of two identical experimental Deltamax reactors composed of 0.001-inch tape. Each wound with two windings; winding A: 5,000 turns of no. 33 wire gage, 145 ohms resistance, used as gate winding of the series-connected saturable reactor; winding B: 5,000 turns of no. 33 wire gage, 190 ohms resistance, used as the control winding of series-connected saturable reactor

Finite Control Resistance. The effect of control resistance on the steady-state behavior of a series-connected saturable reactor with capacitive loading may be seen from the measured relationship of Fig. 3. The relationships given correspond to three values of control resistances $M^2R_c/R_g=2$, 6, and ∞ respectively at a fixed phase angle of 70 degrees. This shows that as the control resistance is reduced, the bistable behavior of the circuit becomes less marked and disappears when the control resistance is sufficiently reduced, such as $M^2R_c/R_g=2$, where the operation is described by a single value curve *OBCEF*.

Further investigation of the circuit for the case of finite control impedance indicated that at some regions of operation where the d-c transformer law holds (indicated by the line *OM* of Fig. 3), the circuit may become regenerative. Thus, at low values of control voltage compared to the referred gate voltage as indicated by the region *OB*, the circuit oscillates at a frequency lower than the carrier frequency. The frequency of oscillation depends on the value of resistance present in the circuit. The oscillatory behavior of the circuit may be predicted from the dynamic analysis of the circuit.

An attempt to extend the steady-state analysis given in reference 1 for the limiting case of infinite control impedance to the more general cases of finite control resistance and dynamic behavior will be extremely laborious if not impossible. This difficulty may be avoided by assuming a rectangular instead of a sinusoidal excitation. The results derived in this paper on the basis of such an assumption confirms the possibility of the existence of a low-frequency oscillation. The frequency of oscillation and the condition for such oscillations is derived. The results derived on the basis of rectangular excitation predict the behavior of the circuit for the case of sinusoidal excitation approximately. This is true since a series connected saturable reactor is essentially a device whose operation may be expressed in terms of half-cycle average values of input and output.

Theory and Assumption

The experimental circuit analyzed consists of two identical reactors with Deltamax* cores of rectangular hysteresis loop. Each core is wound with two equal windings of design details shown in Fig. 4. For a reactor using Deltamax or similar core materials, the magnetization current in many applications may be neglected. Thus for the purpose of analysis, an ideal magnetization characteristic of Fig. 5 may be assumed. On the basis of this assumption the experimental circuit may be

---

* Trade name, Arnold Engineering Company, Marengo, Ill.

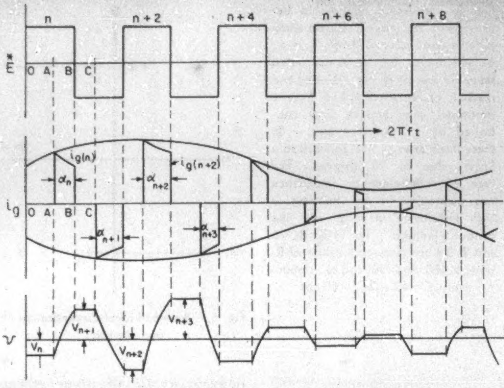


Fig. 6. Waveshape of gate current flowing in idealized series-connected saturable reactor of Fig. 1 and voltage across condenser for rectangular gate excitation and d-c control. Waveshape corresponds to oscillatory case where control voltage is assumed to be below critical value at which sustained oscillation starts

represented by the ideal circuit of Fig. 1, which fulfills the following conditions:

The effective gate resistance R_g is the sum of the two gate-winding resistances and the external gate resistance, $R_g = 2R_{wg} + R_{Lg}$.

The effective control resistance R_c is the sum of the two control-winding resistances and the external control resistance, $R_c = R_{wc} + R_{Lc}$.

The inductance of the core is either zero or infinity.

Hysteresis and eddy-current losses are negligible.

The distributed and winding capacitance of the reactor is negligible, and the leakage of the load condenser is negligible.

Assuming a rectangular gate voltage and a d-c control voltage, the ideal circuit is analyzed. The theoretical results deduced are compared with measured results obtained from the experimental circuit. The correlation between theoretical and measured results confirms the validity of the idealized assumptions made.

Modes of Operation and Circuit Equations

Let a rectangular gate voltage of constant amplitude be applied to the circuit of Fig. 1. In the linear region of operation where d-c transformer law holds as described by line *OM* of Fig. 3, the resulting current pulses and the voltage across the condenser are shown in Fig. 6, where α, V, i_g, and I_g denote the conduction angle, constant voltage across the condenser during nonconduction intervals, instantaneous value of gate current, and the half-cyclic average value of the gate current, respectively. It is assumed that the reader has some basic knowledge regarding the behavior and the mechanics of operation of the circuit for the case of pure resistive loading. This is necessary to understand the subsequent analysis given. For this purpose reading of reference 3, dealing partly with the dynamic behavior of a series-connected saturable reactor with resistive loading, is recommended. For the case of capacitive loading, the mechanics of circuit operation corresponding to the direction of current and voltage indicated in Figs. 1 and 6 is as follows:

During the flow of negative current pulses, reactor I is saturated while reactor II is unsaturated.

During the flow of positive current pulses, reactor I is unsaturated while reactor II is saturated.

During the nonconduction intervals, both reactors are unsaturated.

For these possible modes of operation, it may be concluded that the gate current is a transformation current equal to the value of control current referred to the gate side. The behavior of the circuit may be determined in terms of three unknowns, α, V, and I_g, already defined. Consequently three difference equations are required. These equations may be derived as follows.

Equation I. The first equation may be obtained from consideration of the flux change in core I. At the point O, core I just enters the unsaturated state. It remains in this state until the point *C* is reached where it saturates. Consequently, the net change in flux of core I during the time interval *OABC* is zero. The flux is given by the equation

$$M\frac{d\Phi}{dt} = e_I$$

From this equation it may be concluded that the change in flux is proportional to the volt-time area $\int e_I dt$. Thus during the interval *OABC*, the volt-time area $\int e_I dt = 0$. Since the volt-time area across the reactor during a given interval is proportional to the average voltage across the reactor, consequently the average value of the voltage e_I, across the reactor during the interval *OABC* is zero. Making use of this fact, the difference equation may be obtained as follows.

During the interval *OA*, both reactors are unsaturated, thus it may be shown that

$$e_I = \frac{1}{2}\left[V_n + E^* - ME_c\right] \tag{1}$$

Note: During the intervals when both cores are unsaturated (such as *OA* and *BC* of Fig. 6), the behavior of the circuit is described by the following equations, which may be obtained by applying Kirchhoff's and Ampere's laws to the circuit which gives:

$$e_c = R_c i_c + e_{II}/M - e_I/M \tag{1A}$$

$$e = R_g i_g + v + e_I + e_{II} \tag{1B}$$

$$Mi_g = i_c \tag{1C}$$

$$Mi_g = -i_c \tag{1D}$$

From equations 1(C) and 1(D) it may be seen that $i_c = i_g = 0$. Substituting this condition in equations 1(A) and 1(B) and solving for e_I, gives:

$$e_I = (e - Me_c - V)/2$$

During the interval *OA* (Fig. 6), $e = +E^*$, $e_c = +E$, $v_c = -V_n$, thus

$$e_I = +(V_n + E^* - ME_c)/2$$

During the interval *BC* (Fig. 6), $e = -E^*$, $e_c = +E_c$, $v = +V_{n+1}$, thus

$$e_I = -(V_{n+1} + E^* + ME_c)/2$$

The half-cycle average value of e_1 during interval *OA* is:

$$[e_I]_{OA} = \frac{1}{2}\left[V_n + E^* - ME_c\right]\left[\frac{\pi - \alpha_n}{\pi}\right] \tag{2}$$

Similarly during the interval *BC* both reactors are unsaturated, thus

$$[e_I] = -\frac{1}{2}\left[V_{n+1} + E^* + ME_c\right]$$

(see note following equation 1) **(3)**

The half-cycle average value of e_I during this interval is:

$$[e_I]_{BC}=-\frac{1}{2}[V_{n+1}+E^*+ME_c]\left[\frac{\pi-\alpha_{n+1}}{\pi}\right] \tag{4}$$

During the interval AB, reactor II is saturated while reactor I is unsaturated, thus it may be shown that:

$$-\frac{1}{M}e_I=E_c-R_ci_c$$

or

$$e_I=M^2R_ci_{g(n)}-ME_c$$

The half-cycle average value of e_I during this interval is given by:

$$[e_I]_{AB}=M^2R_cI_{g(n)}-ME_c\frac{\alpha_n}{\pi} \tag{5}$$

where $I_{g(n)}$ is the half-cycle average value of load current during nth half-cycle. As previously deduced,

$$[e_I]_{OA}+[e_I]_{AB}+[e_I]_{BC}=0$$

Substituting equations 2, 4, and 5 in this equation gives the required difference equation which is of the form:

$$[V_n-V_{n+1}-2ME_c]-[V_n+E^*+ME_c]\times\frac{\alpha_n}{\pi}+[V_{n+1}+E^*+ME_c]\frac{\alpha_{n+1}}{\pi}+2\left[\frac{M^2R_c}{R_g}\right]R_gI_{g(n)}=0 \tag{6}$$

Equation II. If Q_n and Q_{n+1} are the charges across the condenser during the nonconduction intervals of nth and $n+1$ half-cycles, then

$$Q_{n+1}=-Q_n+\frac{1}{\omega C}\int_{\pi-\alpha_n}^{\pi}i_{g(n)}d\omega t=-Q_n+\frac{\pi}{\omega C}I_{g(n)}$$

Since $Q_n=CV_n$, then

$$\frac{\pi}{\omega C}I_{g(n)}=V_n+V_{n+1}$$

This equation may be written as:

$$\pi\tan\bar{\phi}\left(1+\frac{M^2R_c}{R_g}\right)R_gI_{g(n)}=V_n+V_{n+1} \tag{7}$$

where

$$\bar{\phi}=\tan^{-1}\frac{1}{\omega C(R_g+M^2R_c)}$$

which is the effective phase angle during the conduction interval.

Equation III. During the nth conduction interval, reactor II is saturated while reactor I is unsaturated. The differential equation of the circuit is of the form:

$$[M^2R_c+R_g]i_{g(n)}+\frac{1}{C}\int i_{g(n)}dt=E^*+ME_c$$

Solution of this equation gives:

$$i_{g(n)}=K\epsilon^{-\frac{t}{R_g+M^2R_g}\frac{1}{C}}=K\epsilon^{-\tan\bar{\phi}\omega t}$$

Substituting the boundary condition

$$i_{g(n)}=\frac{E^*+V_n+ME_c}{(R_g+M^2R_c)}\text{ at }\omega t=\pi-\alpha_n$$

the value of the constant K may be determined. The equation of the current then becomes:

$$i_{g(n)}=\frac{E^*+V_n+ME_c}{(R_g+M^2R_c)}\epsilon^{\tan\bar{\phi}(\pi-\alpha_n)}\epsilon^{-\tan\bar{\phi}\omega t}$$

The half-cycle average value $I_{g(n)}$ of the current may be obtained by integrating this equation between the limits $\omega t=\pi-\alpha_n$ and $\omega t=\pi$ which gives:

$$[1-\epsilon^{-\tan\bar{\phi}\alpha_n}]=\frac{\pi\tan\bar{\phi}}{E^*+V_n+ME_c}\times\left[1+\frac{M^2R_c}{R_g}\right]R_gI_{g(n)} \tag{8}$$

From these three equations both the steady state and the transient behavior of the circuit may be obtained.

Steady-State Behavior

For steady-state operations equations 6, 7, and 8 reduce to

$$-2ME_c+2M^2R_cI_g=0 \tag{9}$$

$$\pi\tan\bar{\phi}\left(1+\frac{M^2R_c}{R_g}\right)R_gI_g=2V \tag{10}$$

$$[1-\epsilon^{-\tan\bar{\phi}\gamma}]=\frac{\pi\tan\bar{\phi}}{E^*+V+ME_c}\left[1+\frac{M^2R_c}{R_g}\right]R_gI_g \tag{11}$$

The first equation may be written as:

$$I_g=\frac{R_c}{MR_c}=\frac{I_c}{M} \tag{12}$$

This equation expresses analytically the fact that the circuit obeys the d-c transformer rule where the half-cycle average value of gate current equals the direct current flowing in the control winding, referred to the gate.

To determine the steady-state relationship between conduction angle and control voltage, equation 11 may be written

$$\epsilon^{-\gamma\tan\bar{\phi}}=\frac{E^*+V+ME_c-\pi\tan\bar{\phi}\left[\frac{M^2R_c}{R_g}+1\right]R_gI_g}{E^*+V+ME_c}$$

Elimination of V from this equation and equation 10 gives:

$$\epsilon^{-\gamma\tan\bar{\phi}}=\frac{E^*-\left[\frac{\pi\tan\bar{\phi}}{2}\left(\frac{M^2R_c}{R_g}+1\right)R_gI_g-ME_c\right]}{E^*+\left[\frac{\pi\tan\bar{\phi}}{2}\left(\frac{M^2R_c}{R_g}+1\right)R_gI_g+ME_c\right]}$$

Substituting the value of I_g given by equation 12 in this equation gives:

$$\epsilon^{-\gamma\tan\phi}=\frac{1-\left[\frac{\pi\tan\bar{\phi}}{2}\left(1+\frac{R_g}{M^2R_c}\right)-1\right]\frac{ME_c}{E^*}}{1+\left[\frac{\pi\tan\bar{\phi}}{2}\left(1+\frac{R_g}{M^2R_c}\right)+1\right]\frac{ME_c}{E^*}}$$

from which

$$\frac{ME_c}{E^*}=\frac{1}{\frac{\pi\tan\bar{\phi}}{2}\left[1+\frac{R_g}{M^2R_c}\right]\left[\frac{\epsilon^{\gamma\tan\bar{\phi}}+1}{\epsilon^{\gamma\tan\bar{\phi}}-1}\right]-1} \tag{13}$$

From this equation the steady-state nondimensional relationship between conduction angle and the control voltage may be calculated.

The steady-state behavior is not always possible since at some ranges of circuit parameters a low-frequency oscillation may occur where the half-cycle average value of current is no longer constant, but modulated by a low-frequency signal as shown in Fig. 6. This behavior may be predicted from the dynamic equations of the circuit as in the following section.

Oscillatory Behavior

Equation 8 contains nonlinear terms. The difficulty of solving this equation may be eliminated by making some justifiable approximations. For small variations about a given operating point, equation 8 may be written as:

$$1-\frac{\epsilon^{\tan\bar{\phi}(\gamma-\alpha_n)}}{\epsilon^{\tan\bar{\phi}\gamma}}=\frac{\pi\tan\bar{\phi}}{ME_c+E^*+V_n}\left[1+\frac{M^2R_c}{R_g}\right]R_gI_{g(n)}$$

where γ is a steady-state operation conduction angle about which a low-frequency oscillation α_n takes place. Thus, for small oscillation corresponding to small values $(\gamma-\alpha_n)$, this equation may be written as:

$$1-\frac{1+\gamma\tan\bar{\phi}-\tan\bar{\phi}\alpha_n}{\epsilon^{\tan\bar{\phi}\gamma}}=\frac{\pi\tan\bar{\phi}}{ME_c+E^*+V_n}\left[1+\frac{M^2R_c}{R_g}\right]R_gI_{g(n)}$$

This equation may be written as:

$$\alpha_n=\left[\frac{R_gI_{g(n)}}{ME_c+E^*+V_n}\right]K_1-K_2 \tag{14}$$

where K_1 and K_2 are constants corresponding to the steady-state conduction angle, and are given by:

$$K_1=\pi\epsilon^{\gamma\tan\bar{\phi}}\left[1+\frac{M^2R_c}{R_g}\right]$$

$$K_2=\frac{\epsilon^{\gamma\tan\bar{\phi}}-\gamma\tan\bar{\phi}-1}{\tan\bar{\phi}} \tag{15}$$

Substituting equation 14 in equation 6 gives:

$$[V_n - V_{n+1} - 2ME_c] - \frac{1}{\pi}[V_n + E^* + ME_c] \times \left[K_1\left(\frac{R_g I_{g(n)}}{ME_c + E^* + V_n}\right) - K_2\right] + \frac{1}{\pi}[V_{n+1} + E^* + ME_c] \times \left[K_1\left(\frac{R_g I_{g(n+1)}}{ME_c + E^* + V_n} + 1\right) - K_2\right] + \frac{2M^2R_c}{R_g} R_g I_{g(n)} = 0$$

from which

$$[V_n - V_{n+1} - 2ME_c] + \frac{K_2}{\pi}[V_n - V_{n+1}] - \frac{K_1}{\pi}[R_g I_{g(n)} - R_g I_{g(n+1)}] + \frac{2M^2R_c}{R_g} R_g I_{g(n)} = 0$$

This equation may be written as:

$$ME_c - \frac{M^2R_c}{R_g} R_g I_{g(n)} + \frac{K_1}{2\pi}[R_g I_{g(n)} - R_g I_{g(n+1)}] - \frac{1}{2}\left[1 + \frac{K_2}{\pi}\right][V_n - V_{n+1}] = 0 \quad (16)$$

The next step is to eliminate the V terms from this equation. This may be done as follows.

Letting $n = n+1$ and $n+1 = n+2$ in equation 16 gives:

$$ME_c - \frac{M^2R_c}{R_g} R_g I_{g(n+1)} + \frac{K_1}{2\pi}[R_g I_{g(n+1)} - R_g I_{g(n+2)}] - \frac{1}{2}\left[1 + \frac{K_2}{\pi}\right] \quad [V_{n+1} - V_{n+2}] = 0 \quad (17)$$

Adding equations 16 and 17 gives:

$$2ME_c - \frac{M^2R_c}{R_g} R_g I_{g(n)} + R_g I_{g(n+1)}] + \frac{K_1}{2\pi}[R_g I_{g(n)} - R_g I_{g(n+2)}] - \frac{1}{2}\left[1 + \frac{K_2}{\pi}\left[[(V_n + V_{n+1}) - (V_{n+1} + V_{n+2})] = 0 \quad (18)\right.\right.$$

Equation 7 may be written as:

$$V_n + V_{n+1} = \pi \tan \bar{\phi}\left(1 + \frac{M^2R_c}{R_g}\right) R_g I_{g(n)} \quad (19)$$

$$V_{n+1} + V_{n+2} = \pi \tan \bar{\phi}\left(1 + \frac{M^2R_c}{R_g}\right) R_g I_{g(n+1)} \quad (20)$$

Thus, substituting equations 19 and 20 in equation 18 gives:

$$R_g I_{g(n+2)} + \frac{\pi}{K_1}\left[\frac{2M^2R_c}{R_g} - \left(1 + \frac{K_2}{\pi}\right) \times \left(1 + \frac{M^2R_c}{R_g}\right)\pi \tan \bar{\phi}\right] R_g I_{g(n+1)} + \frac{\pi}{K_1}\left[\frac{2M^2R_c}{R_g} - \frac{K_1}{\pi} + \left(1 + \frac{K_2}{\pi}\right) \times \left(1 + \frac{M^2R_c}{R_g}\right)\pi \tan \bar{\phi}\right] R_g I_{g(n)} - \frac{4\pi}{K_1} ME_c = 0 \quad (21)$$

Letting

$$I_{g(n+1)} = I_{g(n)} + \Delta I_{g(n)}$$

and

$$I_{g(n+2)} = I_{g(n)} + 2\Delta I_{g(n)} + \Delta^2 I_{g(n)}$$

and arranging similar terms gives:

$$\Delta^2 I_{g(n)} + \left[2 + \frac{\pi}{K_1}\left\{\frac{2M^2R_c}{R_g} - \left(1 + \frac{K_2}{\pi}\right) \times \left(1 + \frac{M^2R_c}{R_g}\right)\pi \tan \bar{\phi}\right\}\right] \Delta I_{g(n)} + \left[\frac{4\pi}{K_1} \frac{M^2R_c}{R_g}\right] I_{g(n)} + \frac{4\pi}{K_1} \frac{ME_c}{R_g} = 0$$

Substituting values of K_1 and K_2 given by equation 15 in this equation gives:

$$\Delta^2 I_{g(n)} + \left[\epsilon^{+\gamma \tan \bar{\phi}} + \gamma \tan \bar{\phi} - \pi \tan \bar{\phi} + 1 + \frac{2}{1 + \frac{R_g}{M^2R_c}}\right] \epsilon^{-\gamma \tan \phi} \Delta I_{g(n)} + \frac{4\epsilon^{-\gamma \tan \phi}}{1 + \frac{R_g}{M^2R_c}}\left[I_{g(n)} - \frac{ME_c}{M^2R_c}\right] = 0$$

Letting $(I_{g(n)} - E_c/MR_c) = I_g{}'_{(n)}$, where $I_g{}'_{(n)}$ is the variation of the half-cycle average value of gate current from its steady-state average value, gives:

$$\Delta^2 I_{g(n)}{}' + B\Delta I_{g(n)}{}' + CI_{g(n)}{}' = 0 \quad (22)$$

where

$$B = \left[\epsilon^{+\gamma \tan \phi} + \gamma \tan \phi - \pi \tan \bar{\phi} + 1 + \frac{2}{1 + \frac{R_g}{M^2R_c}}\right] \epsilon^{-\gamma \tan \phi} \quad (23)$$

$$C = \left[\frac{4\epsilon^{-\gamma \tan \phi}}{1 + \frac{R_g}{M^2R_c}}\right] \quad (24)$$

This is the dynamic equation of the circuit, which is a second-order linear difference equation. For slow variations this difference equation may be approximated by a second-order differential equation of the form:

$$\frac{d^2 I_{g(n)}{}'}{dn^2} + B \frac{dI_{g(n)}{}'}{dn} + CI_{g(n)}{}' = 0 \quad (25)$$

This approximate equation is only valid for slow variations where both the critical condition for sustained oscillation and the frequency of oscillation may be calculated approximately. For more accuracy and faster variations the difference equation must be solved directly.

CRITICAL CONDITION FOR THE EXISTENCE OF LOW-FREQUENCY SUSTAINED OSCILLATION

At high values of control voltage, the circuit of Fig. 1 is stable. As the control is reduced, a critical point is reached where the circuit becomes unstable and oscillatory. The critical condition for sustained oscillation may be obtained from consideration of the stability of equation 22. For this equation to be stable three conditions must be satisfied:

1. $C > 0$
2. $2B - C < 4$
3. $B - C > 0$

From equation 24 it may be seen that C is always positive; that is, the first condition is always satisfied. From an examination of of equations 23 and 24 it may be seen that the second condition is always satisfied. Therefore, the reason for instability is the failure of the third condition. At the critical boundary between stable and unstable regions, $B = C$. Equating values of B and C given by equations 23 and 24 and letting $\gamma = \gamma_c$, gives:

$$\epsilon^{\gamma_c \tan \phi} + \gamma_c \tan \bar{\phi} - \pi \tan \bar{\phi} - \frac{2}{1 + \frac{R_g}{M^2R_c}} + 1 = 0 \quad (26)$$

From this equation the critical value of conduction angle γ_c at which the circuit becomes unstable may be determined. For values of γ below this critical value, the circuit is unstable.

The instability may or may not be associated with oscillatory behavior. The condition for sustained oscillation may be determined by taking the Laplace transform of equation 22, which gives:

$$[I_g{}'(S) - I_{g(0)}{}' - \epsilon^{-S} I_{g(1)}{}']\epsilon^{2S} + [B - 2] \times [I_g{}'(S) - I_{g(0)}{}']\epsilon^{S} + [1 - B + C] I_g{}'(S) = 0$$

Letting $\epsilon^S = Z$, this equation may be written as:

$$I_g{}'(Z) = \frac{[Z^2 + (B-2)Z] I_{g(0)}{}' + Z I_{g(1)}{}'}{Z^2 + [B-2]Z + [1 - B + C]}$$

The behavior of the circuit is oscillatory

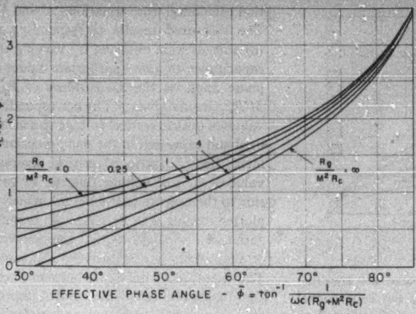


Fig. 7. Plot of $\gamma_c \tan \bar{\phi}$ against $\bar{\phi}$ calculated from equation 27 for series-connected saturable reactor with capacitive loading and finite control resistance for case of rectangular gate excitation and d-c control

when $B^2/4-C<0$. For this case this equation may be written as:

$$I_g'(Z)=\frac{[Z^2+(B-2)Z]I_{g(0)}'+ZI_{g(1)}'}{\left\{Z-\left[\left(1-\frac{B}{4}\right)+j\sqrt{C-\frac{B^2}{4}}\right]\right\}\times\left\{Z-\left[\left(1-\frac{B}{2}\right)-j\sqrt{C-\frac{B^2}{4}}\right]\right\}}$$

Assuming zero initial condition, $I_{g(0)}'=0$, this equation reduces to

$$I_g(Z)=\frac{ZI_{g(1)}'}{\left\{Z-\left[\left(1-\frac{B}{2}\right)+j\sqrt{C-\frac{B^2}{4}}\right]\right\}\times\left\{Z-\left[\left(1-\frac{B}{2}\right)-j\sqrt{C-\frac{B^2}{4}}\right]\right\}}$$

Frequency of Oscillation

Taking the inverse transform of equation 27 gives:

$$I_{g(n)}'=\frac{I_{g(1)}'}{2j\sqrt{C-\frac{B^2}{4}}}\times\left\{\left[\left(1-\frac{B}{2}\right)+j\sqrt{C-\frac{B^2}{4}}\right]^n-\left[\left(1-\frac{B}{2}\right)-j\sqrt{C-\frac{B^2}{4}}\right]^n\right\} \quad (27)$$

This equation may be put into the following form:

$$I_{g(n)}'=\frac{[1-B+C]^{n/2}}{\sqrt{C-\frac{B^2}{4}}}\,I_{g(1)}'\times\sin\left\{\left[\tan^{-1}\frac{\sqrt{C-\frac{B^2}{4}}}{\left(1-\frac{B}{2}\right)}\right]n\right\} \quad (28)$$

From this equation the frequency of oscillation referred to the duration of half a cycle taken as unit-time may be calculated and is given by

$$\frac{1}{2\pi}\tan^{-1}\left[\sqrt{C-\frac{B^2}{4}}\Big/\left(1-\frac{B}{2}\right)\right]$$

The true frequency of oscillation f_m may be obtained by dividing this frequency by $(1/2f)$ which is the time required for half a cycle. Thus

$$\frac{f_m}{f}=\frac{1}{\pi}\tan^{-1}\left[\sqrt{C-\frac{B^2}{4}}\Big/\left(1-\frac{B}{2}\right)\right] \quad (29)$$

From this equation the ratio of modulation to carrier frequency, (f_m/f), may be calculated, where B and C are given by equations 23 and 24.

At the critical point where the system just becomes unstable, $B=C$, thus equation 29 becomes

$$\frac{f_m}{f}=\frac{1}{\pi}\tan^{-1}\left[\sqrt{C-\frac{C^2}{4}}\Big/\left(1-\frac{C}{2}\right)\right]$$

Substituting the value of C given by equation 24 in this equation and letting $\gamma=\gamma_c$, gives:

$$\frac{f_m}{f}=\frac{1}{\pi}\tan^{-1}\left[\frac{\sqrt{\left(1+\frac{R_g}{M^2R_c}\right)\epsilon^{+\gamma_c\tan\bar{\phi}}-1}}{\frac{1}{2}\left(1+\frac{R_g}{M^2R_c}\right)\epsilon^{+\gamma_c\tan\bar{\phi}}-1}\right] \quad (30)$$

This equation gives the ratio of modulation to carrier frequency at the critical point where the circuit becomes oscillatory. The variation of (f_m/f) is smooth, that is, the low frequency oscillation is not necessarily a subharmonic frequency of the carrier.

Discussion of the Theoretical Results

The critical value of conduction angle γ_c at which the circuit becomes oscillatory is given by equation 26. This equation is represented graphically in Fig. 7, which is a plot of $(\gamma_c \tan \bar{\phi})$ against $\bar{\phi}$ for different values of $(R_g/M^2R_c)=0$, 0.25, 1, 4, and ∞, taken as parameters.

Substituting values of γ_c and the corresponding values of $\bar{\phi}$ expressed by the graphical relationship of Fig. 7 in equation 13, the critical value of control voltage corresponding to any value of phase angle may be calculated. This relationship is expressed graphically in Fig. 8, which is a plot of (ME_c/E^*) against $\bar{\phi}$, for different values of $(R_g/M^2R_c)=0$, 0.25, 1, 4, and ∞, taken as parameters.

Similarly substituting values of γ_c and the corresponding values of $\bar{\phi}$ expressed by the graphical relationship of Fig. 7 in equation 30, the ratio of modulation to carrier frequency (f_m/f) may be calculated. This relationship is expressed graphically in Fig. 9 which is a plot of (f_m/f) against

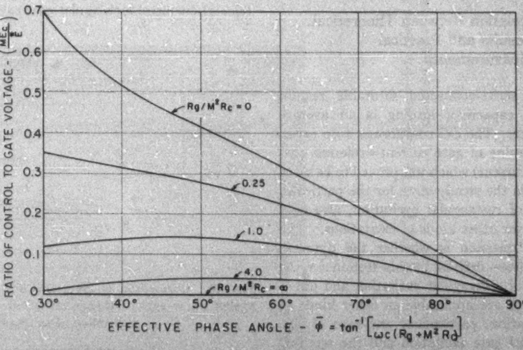


Fig. 8. Plot of ME_c/E^* (ratio of referred control voltage to gate voltage) against $\bar{\phi}$ (effective phase angle), calculated from equations 27 and 13 for series-connected saturable reactor with capacitive loading and finite control resistance for case of rectangular gate excitation and d-c control

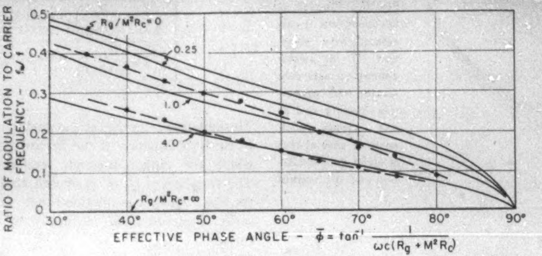


Fig. 9. Plot of f/f_m (ratio of modulation to carrier frequency) against $\bar{\phi}$ calculated from equations 27 and 30 for series-connected saturable reactor with capacitive loading and finite control resistance having rectangular gate excitation and d-c control. Two practical curves indicated on same figures were obtained from tabulated results of Fig. 11

$\bar{\phi}$ for fixed values of $(R_g/M^2R_c) = 0, 0.25, 1, 4$, and ∞ taken as parameters.

From these graphical relationships, the critical value of control voltage below which self-sustained oscillation starts and the frequency of these oscillations may be calculated. It may also be concluded that:

The modulation frequency is always less than the carrier frequency.

For a given value of R_g/M^2R_c, the modulation frequency decreases as the phase angle is increased.

For a given value of phase angle the modulation frequency decreases as R_g/M^2R_c is increased.

For a given value of R_g/M^2R_c the critical value of control voltage increases with the phase angle until it reaches a maximum; it then reduces gradually to zero as the phase angle is further increased.

For a given value of phase angle the critical value of control voltage decreases as R_g/M^2R_c is increased.

Correlation Between Theoretical Results and Practical Measurements

A series-connected saturable reactor with capacitive loading is an average device. The *CC* transformer rule (average value of gate current = referred control current) which was proved to be satisfied in the steady state for the particular case of rectangular excitation, also hold true for other kinds of excitations. The next question is whether the dynamic equations derived in this section for the case of rectangular excitation and direct control are applicable to other kinds of excitation, particularly the case of sinusoidal gate excitation and the rectified sinusoidal control. The answer to this question is that the dynamic behavior of the circuit changes with the condition of excitation; however, the behavior remains qualitatively the same, regardless of the waveshape of the forcing functions. The quantitative difference in the behavior in some ranges of operation is small enough that the dynamic behavior of the circuit is approximately the same regardless of the type of excitation. These conclusions were arrived at from the following investigations of the experimental series-connected saturable reactor circuit using two reactors of Fig. 4.

1. For three fixed values of $R_g/M^2R_c = 4$, 1, and 0.25, values of the critical control voltage corresponding to different values of effective phase angle $\bar{\phi} = \tan^{-1} 1/\omega C(R_g + M^2R_c)$, are plotted in Fig. 10 for these three cases:

(*a*). Rectified a-c control and sinusoidal gate excitation obtained by measurement on the experimental circuit.

(*b*). D-c control and sinusoidal gate excitation obtained by measurement on the experimental circuit.

(*c*). D-c control rectangular gate excitation calculated from equations 26 and 13.

The measured results corresponding to (*a*) and (*b*) were obtained by varying the capacitance to give appropriate values of phase angle at the fixed values of R_g/M^2R_c, already given. The corresponding values of critical control voltage was measured and expressed in the nondimensional form ME_c/E (ratio of critical average value of control voltage referred to gate side to the average gate voltage), which is plotted against $\bar{\phi}$ in Fig. 10. An investigation of Fig. 10 shows that the behavior of the circuit is qualitatively the same. The quantitative difference between these three cases reduces as the control impedance is reduced, such that at a value of $R_g/M^2R_c = 4$, the two measured curves for the case of sinusoidal excitation almost coincide and become closer to the theoretical results obtained by assuming a rectangular excitation.

2. For two fixed values of $R_g/M^2R_c = 4$, and 1, the waveshape of the gate current corresponding to different values of phase angle is given in the oscillograms of Figs. 11(A) and (B), respectively, for the case of sinusoidal gate excitation and d-c control. For each of these oscillograms corresponding to a fixed phase angle $\bar{\phi}$, the control current was adjusted to a critical value at which sinusoidal oscillation started, and the ratio of modulation to carrier frequency f_m/f was obtained by measurements on the oscillogram. Measured values of f_m/f together with the corresponding critical values of control current for different values of phase angle are tabulated in Fig. 11 and are plotted graphically in Fig. 9. From a comparison of these measured results with the theoretical curve of Fig. 9, it may be concluded that the theoretical analysis for the simplified

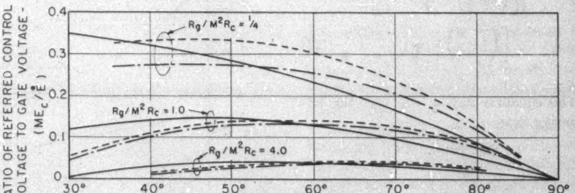
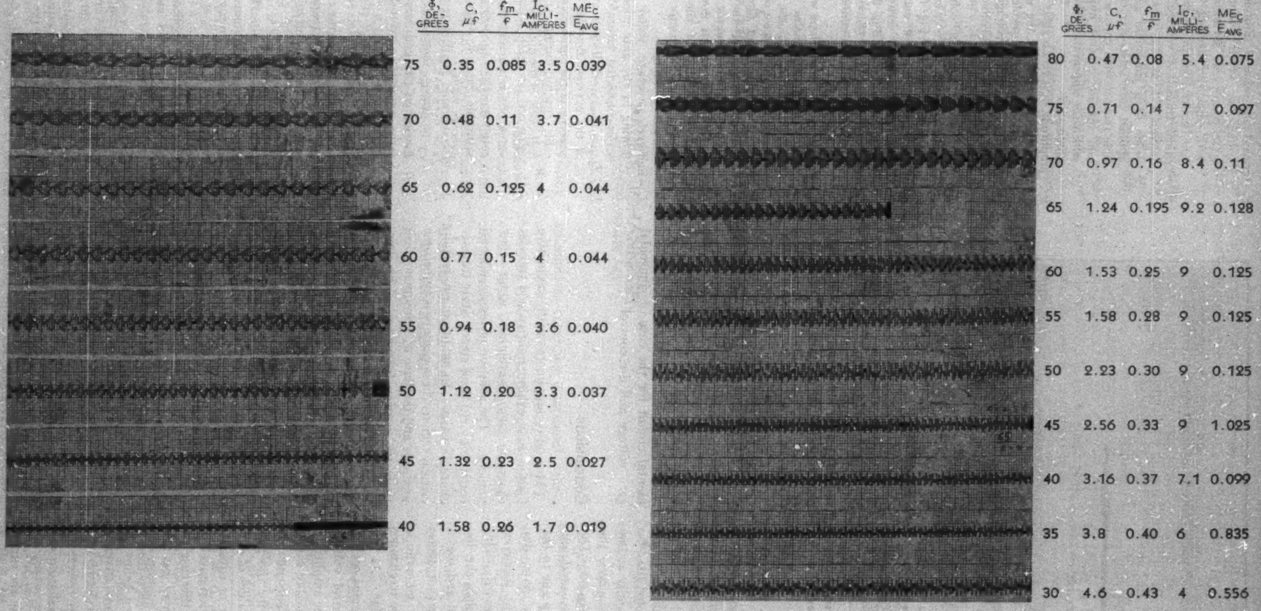


Fig. 10. Comparison of theoretical and measured values of critical control voltage plotted against phase angle. Theoretical results calculated from equations 27 and 13, deduced on assumption of rectangular gate excitation and d-c control

$E^* = 40$ volts rms $= 36$ volts average
$f = 60$ cps
$R_g = 1{,}600$, 500, and 400 ohms
$R_c = 400$, 500, and 1,600 ohms, corresponding to $R_g/M^2R_c = 4$, 1, and 0.25
Value of capacitance was varied to give required phase angle



(A)

Φ, DEGREES	C, μf	$\frac{f_m}{f}$	I_c, MILLIAMPERES	$\frac{ME_c}{E_{AVG}}$
75	0.35	0.085	3.5	0.039
70	0.48	0.11	3.7	0.041
65	0.62	0.125	4	0.044
60	0.77	0.15	4	0.044
55	0.94	0.18	3.6	0.040
50	1.12	0.20	3.3	0.037
45	1.32	0.23	2.5	0.027
40	1.58	0.26	1.7	0.019

(B)

Φ, DEGREES	C, μf	$\frac{f_m}{f}$	I_c, MILLIAMPERES	$\frac{ME_c}{E_{AVG}}$
80	0.47	0.08	5.4	0.075
75	0.71	0.14	7	0.097
70	0.97	0.16	8.4	0.11
65	1.24	0.195	9.2	0.128
60	1.53	0.25	9	0.125
55	1.58	0.28	9	0.125
50	2.23	0.30	9	0.125
45	2.56	0.33	9	1.025
40	3.16	0.37	7.1	0.099
35	3.8	0.40	6	0.835
30	4.6	0.43	4	0.556

Fig. 11. Oscillograms of gate current flowing in series-connected saturable reactor with capacitive loading using two reactors of Fig. 4 for case of sinusoidal gate excitation of 60 cps, and d-c control. Gate voltage = 40 volts rms = 36 average, $f = 60$ cps

A—$R_g/M^2R_c = 4$, $R_g = 1{,}600$ ohms, $R_c = 400$ ohms
B—$R_g/M^2R_c = 1$, $R_g = 500$ ohms, $R_c = 500$ ohms

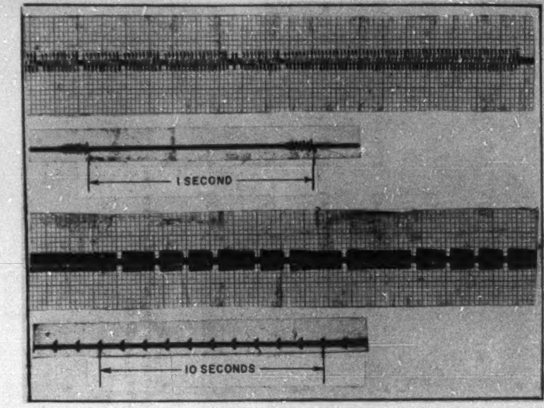


Fig. 12. Irregular oscillation in series-connected saturable reactor with capacitive loading using two reactors of Fig. 4 for case of sinusoidal gate excitation, 60 cps, and d-c control

Gate voltage 40 volts rms
$f=60$ cps
$R_g/M^2R_c=4$
$R_g=1{,}600$ ohms
$R_c=400$ ohms
$0=85$ degrees
$C=0.08/f$

case of the practical circuit for the case of sinusoidal excitation, especially at low values of control resistance where the accuracy of prediction increases. The discrepancy between the measured and the theoretical curves is partly due to the magnetization current. The error due to the magnetization current increases as the phase angle is increased. This reasoning may be used to explain the discrepancy between theoretical and measured results at higher values of phase angle.

The circuit is very sensitive to changes in voltage at the critical point. At this point the circuit suddenly begins oscillating. Thus the critical control voltage may be easily measured. The amplitude, of oscillation increases appreciably with a slight reduction in control current. In the oscillograms of Fig. 11 for the gate current, the corresponding values of control voltage were kept larger or equal to the value where the modulation envelope just touches zero axis. For this case the oscillation is regular and may be predicted from the theoretical analysis. If the control voltage is reduced below the value where the modulation envelope touches the zero axis, the oscillation may become irregular and unpredictable. Such an irregular mode of oscillation may be seen from the oscillograms of Fig. 12, for the gate current flowing in the experimental circuit for the case of sinusoidal gate voltage and d-c control.

Conclusions

Magnetic amplifiers with capacitive loading may exhibit multivalued transfer characteristics containing jump phenomena. The solution of transcendental equations yield these steady-state transfer characteristics. In addition, in certain critical regions these circuits may become unstable and oscillate at a frequency less than the carrier frequency. This low-frequency oscillation acts to modulate the carrier frequency output. The use of difference equations and Z-transform techniques allows the prediction of the critical boundaries of the region in which these oscillations occur, and also the frequency of the oscillations.

Appendix. Nomenclature

A = effective core area, measured area × stacking factor of the core material
B_c = critical value of flux density at which the core saturates
C = load capacitance, farads
E = rms value of sinusoidal voltage applied to the gate of the series-connected saturable reactor
E_c = amplitude of the d-c control voltage of a series-connected saturable reactor
E_e = critical rms value of sinusoidal applied voltage at which extinction in a series-connected saturable reactor with capacitive loading takes place when control circuit is open
E_t' = critical transition voltage defined in reference 1 as the maximum rms value of a sinusoidal gate voltage applied to a series-connected saturable reactor with capacitive loading and infinite control impedance at which the appearance of mode IV is eliminated
E^* = amplitude of a rectangular alternating voltage applied to the gate of the series-connected saturable reactor
e = instantaneous value of voltage applied to the gate of the series-connected saturable reactor
e_I = instantaneous value of voltage across reactor I of the series-connected saturable reactor of Fig. 1
e_{II} = instantaneous value of voltage across reactor II of the series-connected saturable reactor of Fig. 2(A)
f = frequency of the gate (carrier) voltage applied to the series-connected saturable reactor
f_m = frequency of the envelope which modulates the average value of the gate current flowing in the series-connected saturable reactor of Fig. 6
I_c = value of direct current flowing in the control winding of the series-connected saturable reactor
$\bar{I}_c$ = value of direct current flowing in the control winding of the series-connected saturable reactor referred to the gate winding
I_g = half-cycle average value of current flowing in the gate winding of the series-connected saturable reactor
$I_{g(n)}, I_{g(n+1)}, I_{g(n+2)}$ = half-cyclic average value of the current flowing in the gate winding of the series-connected saturable reactor during n, $n+1$, $n+2$ half-cycle of the applied voltage
I_g' = variation of the half-cyclic-average value of the current flowing in the gate winding of the series connected-saturable reactor, from the mean average value
$I_g'_{(n)}, I_g'_{(n+1)}, I_g'_{(n+2)}$ = variation of the half-cyclic-average value of the current flowing in the gate winding of the series-connected saturable reactor during n, $n+1$, and $n+2$ half-cycle of the applied voltage, from the mean average value
i_c = instantaneous value of the current flowing in the control winding of the series-connected saturable reactor
i_g = instantaneous value of gate current flowing in the gate winding of the series-connected saturable reactor
$i_{g(n)}, i_{g(n+1)}, i_{g(n+2)}$ = instantaneous value of the gate current flowing in the gate winding of a series-connected saturable reactor during n, $n+1$, and $n+2$ half-cycles of the applied voltage
M = ratio of gate-to-control winding of the series-connected saturable reactor (N_g/N_c)
N_c = number of turns-per-control winding of the series-connected saturable reactor
N_g = number of turns-per-gate winding of the series-connected saturable reactor
Q = constant condenser charge during non-conduction intervals

Q_n, Q_{n+1}, Q_{n+2} = constant charge across the condenser during the nonconduction intervals of n, $n+1$, and $n+2$ half-cycle of the applied voltage
q = instantaneous value of condenser charge
q_n, q_{n+1}, q_{n+2} = instantaneous value of condenser charge during n, $n+1$, and $n+2$ half-cycle of the applied voltage
R_c = effective control resistance of the series-connected saturable reactor, $R_c = 2R_{wc}+R_{Lc}$
R_g = effective gate resistance of the series-connected saturable reactor, $R_g = 2R_{wg}+R_{Lg}$
R_{wc} = resistance of the control winding of the series-connected saturable reactor
R_{wg} = resistance of the gate winding of the series-connected saturable reactor
V = constant condenser voltage during nonconduction intervals
V_n, V_{n+1}, V_{n+2} = constant voltage across the condenser during the nonconduction intervals of n, $n+1$, and $n+2$ half-cycle of the applied voltage
v = instantaneous condenser voltage
β = instantaneous value of flux density
γ = steady-state operation, or mean conduction angle, during half a cycle of the a-c rectangular voltage applied to the gate winding of the series-connected saturable reactor with finite control resistance, during which current flows, and about which transient variation, or oscillation is assumed to take place
γ_c = critical value of the steady-state operation conduction angle below which the series-connected saturable reactor with capacitive loading and finite control impedance becomes unstable and oscillatory
Φ = total flux
$\bar{\phi}$ = effective phase angle of the series-connected saturable reactor of Fig. 1, $\bar{\phi} = \tan^{-1} 1/\omega C(R_g+M^2R_c)$, degrees or radians
ω = angular frequency of the gate (carrier) voltage applied to the series-connected saturable reactor
ω_m = angular frequency of the envelope which modulates the average value of the gate current flowing in the series-connected saturable reactor, Fig. 6.

References

1. Capacitive Loadng of Nonlinear Magnetic Circuits, J. T. Salihi. *Ph.D. Thesis*, University of California, Berkeley, Calif., 1958.

2. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955.

3. Analysis of Magnetic Amplifiers by the Use of Difference Equations, P. R. Johannessen. *AIEE Transactions*, pt. I, (*Communication and Electronics*), vol. 73, 1954 (Jan. 1955 section). pp. 700–11.

4. Transients in Linear Systems (book), Murray F. Gardner, John L. Barnes. John Wiley & Sons, Inc., 1948.

5. Series-Connected Saturable Reactor with Control Source of Comparatively High Impedance, H. F. Storm. *AIEE Transaction*, vol. 69, pt. II, 1950, pp. 1299–1309.

6. Analysis of Instability and Response of Reactors with Rectangular Hysteresis Loop Core Material in Series with Capacitors, Jalal T. Salihi. *Ibid.*, pt. I (*Communication and Electronics*), vol. 75, July 1956, pp. 296–307.

Analysis of Magnetic Amplifiers Without Diodes

PAUL R. JOHANNESSEN
ASSOCIATE MEMBER AIEE

ALTHOUGH THERE ARE many methods and procedures for analyzing complex circuits, probably the best-known general approach is to apply an organized, systematic procedure to the combination of the analytic expressions for the voltage-current characteristics of the individual circuit elements. To be able to apply this method of attack, it is essential that the characteristics of the individual circuit elements, if such elements can be defined, can be represented by analytic expressions, and that an organized method exists where the analytic representations of the individual circuit elements can be combined to yield the desired result.

In this paper the foundation for such an analysis will be established for magnetic amplifiers without diodes. The element to be considered as the basic magnetic-amplifier element consists of a 2-winding saturable inductor, one winding being in series with an alternating voltage source. Both the saturable inductor and the alternating voltage source are fundamental to the magnetic amplifier, because either element without the other cannot provide power amplification.[1] By proper choice of variables and the use of certain assumptions regarding component characteristics and voltage waveforms, it will be shown that the analytic representation of the basic magnetic-amplifier element can be made linear.

Analysis of Basic Magnetic-Amplifier Element

The basic magnetic-amplifier element, shown in Fig. 1(A), consists of a saturable inductor and an alternating-voltage source. There are two windings on the inductor: the gate winding in series with the source and port 2; and the control winding in series with port 1. In the analysis of the basic element, the inductor characteristics are considered to be ideal and the two inductor windings are assumed to be perfectly coupled. To simplify the analytic expressions obtained, a one-to-one turns ratio of the two windings is also assumed. A different turns ratio can easily be taken into account by the use of an ideal transformer. Because of the assumptions of ideal coupling and one-to-one turns ratio, the inductor can be considered as a 1-winding inductor that is either a short circuit or a open circuit, depending upon the flux level in the inductor core. In fact, the two windings are introduced merely to make possible the interconnections of basic elements to be considered later; at present, the inductor can be thought of to advantage as a 1-winding inductor.

The analysis assumes, initially, arbitrary waveforms for the supply and port voltages. In the course of the analysis the restrictions upon the waveforms that lead to a linear analysis will be established. At this stage the only restriction imposed upon the supply voltage is that its magnitude be limited so that the volt-time integrals of the positive and negative parts are less than the volt-time integral required to drive the core from positive to negative saturation. This requirement can be stated analytically as follows:

$$\int_{t_a}^{t_b} e_s dt < 2n_s AB_s \qquad (1)$$

where

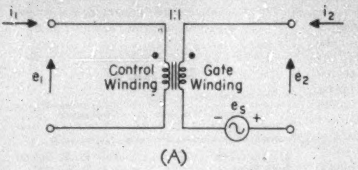
e_s = source voltage
t_a = time of the beginning of a positive or negative half period of e_s

---

Paper 59-171, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted August 15, 1958; made available for printing January 12, 1959.

Paul R. Johannessen is with the Applied Research Laboratory of the Sylvania Electronics System, Waltham, Mass.

The material presented in this paper was submitted to the Massachusetts Institute of Technology, Cambridge, Mass. in partial fulfillment of the requirements for the Doctor of Science degree. It was supported by the U. S. Air Force under Contract No. *AF 33(616)-5489*, MIT Project No. *DSR 7848*.



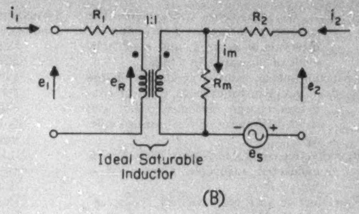


Fig. 1. A—Basic magnetic-amplifier element. B—The modified element

t_b = time of the ending of the same half period
n_2 = number of turns on the gate winding
A = cross section of the core
B_s = saturation flux density of the core

To avoid difficulties that are usually associated with the analysis when zero and infinite impedances or admittances occur, the resistors R_1, R_2, and R_m are added to the basic circuit as shown in Fig. 1(B). The resistors R_1 and R_2 will later be considered to represent the winding resistances, and the resistor R_m to represent the magnetizing impedance of the inductor. At present, however, these resistors are introduced only for convenience. The characteristics of the inductor in Fig. 1(B) are shown in Fig. 2. The scales of the plot are chosen in such a way that the usual B versus H curve is converted into an $\int e_R dt$ versus i_R curve e_R being the ideal saturable-inductor voltage and i_R the inductor current.

To simplify the analysis a fictitious voltage e_R' is introduced. The voltage e_R' is defined as the voltage that would appear across the inductor if the inductor were open-circuited, and it is given by the following expression:

$$e_R' = ae_1 + b(e_s + e_2) \tag{2}$$

where

a = voltage transfer ratio between the e_1 and the e_R terminals, measured with $e_2 = e_s = 0$
b = voltage transfer ratio between the e_s, or e_2, and the e_R terminals, measured with e_2 (or e_s) $= e_1 = 0$

In terms of the circuit parameters, expressions for a and b are as follows:

$$a = \frac{R_m R_2/(R_m + R_2)}{R_1 + R_m R_2/(R_m + R_2)} \tag{3}$$

$$b = \frac{R_m R_1/(R_m + R_1)}{R_2 + R_m R_1/(R_m + R_1)} = \frac{R_1}{R_2} a \tag{4}$$

During the unsaturated period e_R' is the voltage across the inductor. By Thevenin's theorem it is also seen that if R is the resistance faced by the inductor, e_R'/R is the current in the inductor during the saturated period. The component ae_1 of e_R', together with the *negative* of the component $b(e_s+e_2)$, is plotted in Fig. 3. For reference polarities as in Fig. 1(B), the inductor current e_R'/R must be positive during saturation on the upper branch of the inductor characteristic. Thus e_R' must be positive,

$$ae_1 > -b(e_s + e_2) \tag{5}$$

that is, saturation on the upper branch of the inductor characteristic can occur only when the voltage component ae_1 is above the voltage component $-b(e_s+e_2)$ in Fig. 3. Particular attention should be given to the intersections of the two voltage-component curves, because at these instants, labeled t_1, t_2, and t_1' in Fig. 3, both the inductor voltage and the inductor current are zero. If the inductor saturates on the upper branch of the characteristic, saturation must end at the intersections t_1 and t_1'. Similarly, if the inductor saturates on the lower branch of its characteristics, saturation must end at the intersection t_2.

Suppose that at time t_0 the inductor is saturated and that the operating point P on the inductor characteristic is at point P_0 in Fig. 2. Saturation continues and the inductor voltage e_R is zero until the intersection at time t_1 of the component voltages. During the time interval t_0 to t_1, the operating point P moves from P_0 to P_1. After t_1, the inductor is unsaturated; the inductor voltage e_R is equal to the fictitious voltage e_R' which is negative; and the operating point P moves from P_1 down to P_2 during the interval t_1 to t_2. After t_2, the inductor voltage is positive and the operating point P moves back up to P_1 in the interval t_2 to t_3. Because the flux change from P_1 to P_2 is equal to the negative of the flux change from P_2 to P_1, the integral of the inductor voltage during the intervals t_1 to t_2 and t_2 to t_3 must satisfy the equation

$$\int_{t_2}^{t_3} e_R dt = -\int_{t_1}^{t_2} e_R dt \tag{6}$$

or the two areas shaded in Fig. 3 must be equal. At t_3 the inductor saturates; the voltage e_R drops to zero; and saturation continues until t_1', a point in the next period of operation which corresponds to the point t_1 in the first period considered.

The important conclusions from the foregoing considerations are that during the saturated period the inductor voltage is zero and that during the unsaturated period the integral of the inductor voltage is zero; or

$$\int_{t_1}^{t_3} e_R dt = 0 \tag{7}$$

Therefore, during the entire period from

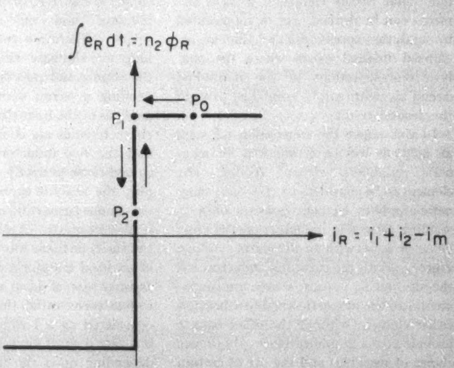


Fig. 2. Ideal saturable inductor

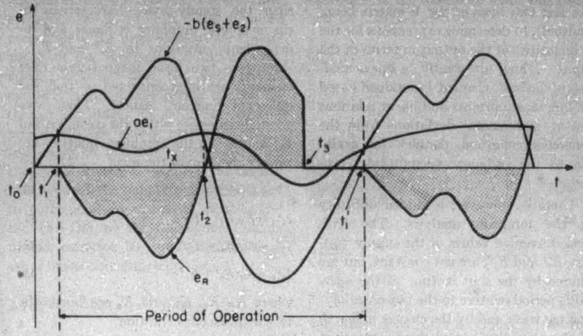


Fig. 3. Voltage waveforms

t_1 to t_1' called the operating period, the integral of the inductor voltage e_R is zero.

The only conditions of the voltage waveforms thus far are:

1. Saturation must occur periodically so that a succession of periods like t_1 to t_1' occur.
2. Saturation must occur at only one end of the inductor characteristic.

For convenience in the remainder of this analysis the following conditions are added:

3. The voltage waveforms must be such that exactly one saturated and one unsaturated period occur in each operating period.
4. The average value of the supply voltage over any complete supply-voltage period must be zero.

Consider now any instant of time t_x within the period t_1 to t_1'. Time t_x divides the operating period into two parts, t_1 to t_x and t_x to t_1'; the first of these periods is called the control period and the second the gating period.

The following relation can now be written

$$\int_{t_1}^{t_x} e_R dt+\int_{t_x}^{t_1'} e_R dt=0 \tag{8}$$

Therefore

$$\frac{t_x-t_1}{t_x-t_1}\int_{t_1}^{t_x} e_R dt+\frac{t_1'-t_x}{t_1'-t_x}\int_{t_x}^{t_1'} e_R dt=0 \tag{9}$$

or

$$(t_x-t_1)E_R^c+(t_1'-t_x)E_R^g=0 \tag{10}$$

where

E_R^c = average value of e_R in the control period

E_R^g = average value of e_R in the gating period

If time t_x divides the interval from t_1 to t_1' into two equal parts so that

$$t_x-t_1=t_1'-t_x \tag{11}$$

the basic relation of equation 10 simplifies to

$$E_R^c+E_R^g=0 \tag{12}$$

It should be realized that the choice of t_x is entirely arbitrary, and the center position is preferred merely because it simplifies the result of the analysis. It will be shown later that there are no compensating disadvantages to this choice for practical circuits. It is also of importance to realize that time t_x does not correspond to time t_3, the instant of time at which the inductor saturates. As a consequence, the control and gating periods do not correspond to the unsaturated and saturated periods. Furthermore, there is no special relation between t_x and t_3; that is, the interval of time during which e_R is negative can be more or less than half the t_1 to t_1' period.

If port 2 is short-circuited and zero signal is applied to port 1, equation 2 reduces to

$$e_R'=be_s \tag{13}$$

Since the voltage e_s by definition does not contain a direct voltage component, the inductor does not saturate; that is, in Fig. 3, t_3 coincides with t_1'. If a positive voltage is applied to port 1, t_3 moves back from t_1' toward t_1. Thus, with small signals applied, t_3 may be expected to occur after t_x, that is in the gating period; only with very large applied signals at port 1 can t_3 occur in the control period. The analysis of interest is concerned primarily with relatively small signals. Also, in practical magnetic amplifiers it is unusual and undesirable for the saturation period, t_3 to t_1', to extend into the control period. It is, therefore, here assumed that t_3 remains in the gating period. This assumption renders the basic circuit linear throughout the control period so that voltage-current relations can be calculated easily.

At any instant at all, the following equations hold

$$e_1=R_1i_1+e_R \tag{14}$$

$$e_2+e_s=R_2i_2+e_R \tag{15}$$

In addition, at any instant in the control period,

$$e_R=ae_1+b(e_s+e_2) \tag{16}$$

These equations completely determine the relations between port quantities during the control period so that any of the 2-port matrices may be calculated. For example, the equations may be solved to yield the admittance-matrix equation:

$$\left\|\begin{matrix} i_1 \\ i_2 \end{matrix}\right\| = \left\|\begin{matrix} \frac{1-a}{R_1} & \frac{-b}{R_1} & \frac{-b}{R_1} \\ \frac{-a}{R_2} & \frac{1-b}{R_2} & \frac{1-b}{R_2} \end{matrix}\right\| \left\|\begin{matrix} e_1 \\ e_2 \\ e_s \end{matrix}\right\| \tag{17}$$

Since these equations are true at each instant, they are also true for average values during the control period.

$$\left\|\begin{matrix} I_1^c \\ I_2^c \end{matrix}\right\| = \left\|\begin{matrix} \frac{1-a}{R_1} & \frac{-b}{R_1} & \frac{-b}{R_1} \\ \frac{-a}{R_2} & \frac{1-b}{R_2} & \frac{1-b}{R_2} \end{matrix}\right\| \left\|\begin{matrix} E_1^c \\ E_2^c \\ E_s^c \end{matrix}\right\| \tag{18}$$

Here, and in subsequent equations, the notation introduced in equation 10 is used; capital letters with superscript *c* are used to denote control-period average values of the quantity denoted by the corresponding lower-case letters. Superscript *g* is used to denote the gating-period average value.

Instantaneous-value equations for the gating period are more difficult to obtain, and a different approach, therefore, is used to derive average-value equations for this period. Equations 14 and 15 are valid at any instant, and in terms of average values in the gating period become

$$E_1^g=R_1I_1^g+E_R^g \tag{19}$$

$$E_2^g+E_s^g=R_2I_2^g+E_R^g \tag{20}$$

By the use of the basic relation

$$E_R^c+E_R^g=0 \tag{12}$$

together with equations 16, 18, 19, and 20, equations analogous to equation 18 can be obtained; and these with equation 18 may be written as the admittance-matrix equation of the network:

$$\left\Vert\begin{matrix} I_1^c \\ \\ I_1^g \\ \\ I_2^c \\ \\ I_2^g \end{matrix}\right\Vert = \left\Vert\begin{matrix} \frac{1-a}{R_1} & 0 & \frac{-b}{R_1} & 0 & \frac{-b}{R_1} & 0 \\ \\ \frac{a}{R_1} & \frac{1}{R_1} & \frac{b}{R_1} & 0 & \frac{b}{R_1} & 0 \\ \\ \frac{-a}{R_2} & 0 & \frac{1-b}{R_2} & 0 & \frac{1-b}{R_2} & 0 \\ \\ \frac{a}{R_2} & 0 & \frac{b}{R_2} & \frac{1}{R_2} & \frac{b}{R_2} & \frac{1}{R_2} \end{matrix}\right\Vert \left\Vert\begin{matrix} E_1^c \\ E_1^g \\ E_2^c \\ E_2^g \\ E_s^c \\ E_s^g \end{matrix}\right\Vert \tag{21}$$

More generally, any desired network matrix can be obtained from these same equations.

Equation 21 appears to indicate that in terms of half-period average values, the basic element can be represented as a linear network containing a source, E_s. Thus, if short circuits are placed at input and output, the following quiescent currents flow:

In the control period,

$${I_1^c}_q = -\frac{b}{R_1}\, E_s^c \tag{22}$$

$${I_2^c}_q = \frac{1-b}{R_2}\, E_s^c \tag{23}$$

In the gating period,

$${I_1^g}_q = \frac{b}{R_1}\, E_s^c \tag{24}$$

$${I_2^g}_q = \frac{b}{R_2}\, E_s^c + \frac{1}{R_2}\, E_s^g \tag{25}$$

As is shown by equations 22 through 25 only alternating currents flow during quiescent condition, inasmuch as E_s^c is equal to $-E_s^g$ in steady-state operation. This result is obvious because no direct-voltage sources are present in the circuit when both network ports are short-circuited.

If now a signal is applied at port 1 with 2 left short-circuited, additional components of currents flow as follows:

$$\Delta I_1^c = \frac{1-a}{R_1}\, E_1^c \tag{26}$$

$$\Delta I_2^c = -\frac{a}{R_2}\, E_1^c \tag{27}$$

$$\Delta I_1^g = \frac{a}{R_1}\, E_1^c + \frac{1}{R_1}\, E_1^c \tag{28}$$

$$\Delta I_2^c = \frac{a}{R_2}\, E_1^c \tag{29}$$

If port 2 is opened and a load applied, the characteristics of the load determine the voltage and current relations at port 2. If the load characteristics can be formulated as a linear relationship among E_2^c, E_2^g, I_2^c, and I_2^g, then the equations stating this relationship can be used with the matrix equation, equation 21 (the last two columns of the Y-matrix and the last two rows of the E-matrix being omitted), to determine expressions for the Δ quantities at the output in terms of the input. Thus, apparently, a linear analysis of the basic element is obtained based on quiescent currents and linear relations among incremental deviations from the quiescent condition, parallel, for example, to the ordinary vacuum-tube procedure.

There is, however, a hidden difficulty in the foregoing analysis. The half-period average values of the supply voltage, E_s^c and E_s^g, are not constant, but are altered by the shift in time of the operating period relative to the power supply-voltage wave and by the change in length of the operating period during transient conditions. This property is seen from Fig. 3, because both t_1 and t_1' are determined by the intersections of the two waves, $-b(e_s+e_2)$ and ae_1, representing components of the fictitious voltage e_R'. If both e_1 and e_2 are zero, t_1 and t_1' are located at the zero crossings of the e_s wave. Existence of the voltages e_1 and e_2 tends to shift t_1 and t_1' from these zero crossings. Thus, a nonlinearity in the basic element is hidden in the terms involving E_s^c and E_s^g and their dependency on the signal.

During steady-state conditions the period of operation t_1 to t_1' is equal to the period T of the power-supply voltage wave so that the nonlinearity in the basic element during steady-state conditions is caused by the shift in the period of operation relative to the power-supply voltage wave only. For practical magnetic amplifiers consisting of interconnections of basic elements, this nonlinearity is usually canceled. For this cancellation to occur, the terms involving E_s^c and E_s^g in the equations representing the operation of a practical magnetic amplifier must cancel one another. Fortunately, conditions under which the terms involving E_s^c and E_s^g cancel can be determined as follows. Equation 21 is a set of linear relations among the half-period average values of the supply voltage and port voltages and currents. It should be realized that in the derivation of these relations the restriction that the supply voltage should remain constant was not imposed. Superposition can therefore be used to determine the effect of variations in the quantities E_s^c and E_s^g upon the magnetic-amplifier operation. It is immaterial what causes the quantities E_s^c and E_s^g to vary. Thus, if it can be shown that the changes in E_s^c and E_s^g, caused by the shift in the period of operation relative to the period of the supply voltage wave, do not affect the quiescent operating conditions, then the steady-state characteristic of the magnetic amplifier is linear and independent variations in E_s^c and E_s^g.

Suppose two basic magnetic-amplifier elements are interconnected so that the quiescent current cancel. Then two possibilities exist; either the coefficients of E_s^c and E_s^g in the amplifier equations are zero or they are of the form

$${I_1^c}_q = K_1(E_s^c + E_s^g) \tag{30}$$

$${I_1^g}_q = K_2(E_s^c + E_s^g) \tag{31}$$

$${I_2^c}_q = K_3(E_s^c + E_s^g) \tag{32}$$

$${I_2^g}_q = K_4(E_s^c + E_s^g) \tag{33}$$

where K_1, K_2, K_3, and K_4 are constants. In the quiescent condition

$$E_s^c = -E_s^g \tag{34}$$

so that the desired quiescent-current cancellation is ensured. During steady-state operation the changes in E_s^c and E_s^g must satisfy the relation

$$\Delta E_s^c + \Delta E_s^g = 0 \tag{35}$$

because the supply voltage contains no direct voltage component and is periodic. Changes in E_s^c and E_s^g during steady-state operation, therefore, do not cause any changes in the short-circuit port currents. Thus, it follows that magnetic amplifiers whose quiescent currents are zero have linear steady-state characteristics independent of variations in E_s^c and E_s^g.

For a single basic element the error introduced by ignoring the variation of the E_s^c and E_s^g terms is not so great as might appear, as is demonstrated in reference 2.

Linear Analytic Representation of the Basic Element

A linear analytic representation of the basic element in terms of incremental half-period average quantities is obtained by eliminating the last two columns in the admittance matrix and the last two rows in the voltage matrix of equation 21, and adding a Δ to the current and voltage symbols. For convenience the added symbol Δ is omitted in writing the following equation.

$$\left\Vert\begin{matrix} I_1^c \\ \\ I_1^g \\ \\ I_2^c \\ \\ I_2^g \end{matrix}\right\Vert = \left\Vert\begin{matrix} \frac{1-a}{R_1} & 0 & \frac{-b}{R_1} & 0 \\ \\ \frac{a}{R_1} & \frac{1}{R_1} & \frac{b}{R_1} & 0 \\ \\ \frac{-a}{R_2} & 0 & \frac{1-b}{R_2} & 0 \\ \\ \frac{a}{R_2} & 0 & \frac{b}{R_2} & \frac{1}{R_2} \end{matrix}\right\Vert \left\Vert\begin{matrix} E_1^c \\ \\ E_1^g \\ \\ E_2^c \\ \\ E_2^g \end{matrix}\right\Vert \tag{36}$$

In the analysis of magnetic amplifiers

Vol.
Pt. 1
1959

not only the static but also the dynamic properties are of interest. The dynamic properties of the basic element are apparent from the superscript notation used, which indicates that the short-circuit currents in the gating period depend on the port voltages in both the control period and the gating period. Thus, a time-delay of one-half period exists, for example, between E_1^c and I_2^g. If the *n*th half cycle of the supply voltage is taken to correspond to a control period, and the $(n+1)th$ to the subsequent gating period, equation 36 can be written as a set of linear difference equations.

$$I_1^c(n)=\frac{1-a}{R_1}E_1^c(n)-\frac{b}{R_1}E_2^c(n)$$

$$I_1^g(n+1)=\frac{a}{R_1}E_1^c(n)+\frac{1}{R_1}E_1^g(n+1)+\frac{b}{R_1}E_2^c(n) \qquad (37)$$

$$I_2^c(n)=-\frac{a}{R_1}E_1^c(n)+\frac{1-b}{R_2}E_2^c(n)$$

$$I_2^g(n+1)=\frac{a}{R_2}E_1^c(n)+\frac{b}{R_2}E_2^c(n)+\frac{1}{R_2}E_2^g(n+1)$$

For the analytic representation of the basic element to serve as a means for analysis of magnetic amplifiers, a more convenient representation than offered by equation 37 is desirable. In linear-circuit analysis such a representation is usually obtained by a transformation from the time domain to the frequency domain, accomplished by either the Fourier or Laplace transform, because such a transformation reduces the mathematical manipulation to simple algebra. To apply this method, however, it is necessary that the variable quantities be transformable. Clearly, the Laplace transform of an average-value quantity has no meaning. Also, there exists no fixed relation between instantaneous voltages and currents of a magnetic amplifier, because the waveforms of the dependent variables depend not only upon the waveform of the independent variables but also upon the waveform of the supply voltage. On an instantaneous basis, therefore, the Laplace transform does not exist. However, for the purpose of analysis, any waveform can be assigned to the variable quantities in equation 37; and, to permit the use of the Laplace-transform method, it is assumed that the variable quantities consist of impulses whose areas are equal to the magnitude of the average-value quantities. This choice of wave form is by no means the only possible one, but is made for convenience. The Laplace transform of the difference equations can now be obtained and yields

$$\left\|\begin{matrix} I_1^c(s) \\ I_1^g(s) \\ I_2^c(s) \\ I_2^g(s) \end{matrix}\right\| = \left\|\begin{matrix} \frac{1-a}{R_1} & 0 & \frac{-b}{R_1} & 0 \\ \frac{a}{R_1}\cdot\frac{1}{z} & \frac{1}{R_1} & \frac{b}{R_1}\cdot\frac{1}{z} & 0 \\ \frac{-a}{R_2} & 0 & \frac{1-b}{R_2} & 0 \\ \frac{a}{R_2}\cdot\frac{1}{z} & 0 & \frac{b}{R_2}\cdot\frac{1}{z} & \frac{1}{R_2} \end{matrix}\right\| \left\|\begin{matrix} E_1^c(s) \\ E_2^c(s) \\ E_1^g(s) \\ E_2^g(s) \end{matrix}\right\| \qquad (38)$$

where

$$z=e^{\frac{T}{2}s} \qquad (39)$$

and where a bar over a symbol is used to denote the Laplace transform of the time function represented by the symbol itself.

The elements of the admittance matrix in equation 38 are the short-circuit driving point and transfer admittances of the basic magnetic-amplifier element. The admittance matrix and its elements may be written

$$Y=\left\|\begin{matrix} y_{11}^{cc} & y_{11}^{cg} & y_{12}^{cc} & y_{12}^{cg} \\ y_{11}^{gc} & y_{11}^{gg} & y_{12}^{gc} & y_{12}^{gg} \\ y_{21}^{cc} & y_{21}^{cg} & y_{22}^{cc} & y_{22}^{cg} \\ y_{21}^{gc} & y_{21}^{gg} & y_{22}^{gc} & y_{22}^{gg} \end{matrix}\right\| \qquad (40)$$

Because impulses are chosen as the waveform of the variable quantities, these admittances can be interpreted simply as the Laplace transform of the unit impulse response. For example, the inverse transform of the y_{21}^{gc} transfer admittance is the short-circuit current at port 2 during the gating periods caused by a voltage applied at port 1 during the first control period (the control period from time 0 to $T/2$).

The admittance matrix of equation 40 can be partitioned into four 2×2 submatrices by writing

$$Y_{rs}=\left\|\begin{matrix} y_{rs}^{cc} & y_{rs}^{cg} \\ y_{rs}^{gc} & y_{rs}^{gg} \end{matrix}\right\| \qquad (41)$$

where *r* and *s* may each be 1 or 2. Thus, the matrix of equation 40 can be written

$$Y=\left\|\begin{matrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{matrix}\right\| \qquad (42)$$

The reader should note that the matrix elements in equation 42 represent submatrices of the form shown by equation 41.

Other relations between port voltages and currents useful in magnetic-amplifier analysis are the following:

$$\left\|\begin{matrix} E_1^c \\ E_1^g \\ E_2^c \\ E_2^g \end{matrix}\right\| = Z \left\|\begin{matrix} I_1^c \\ I_1^g \\ I_2^c \\ I_2^g \end{matrix}\right\| \qquad (43)$$

$$\left\|\begin{matrix} E_1^c \\ E_1^g \\ I_2^c \\ I_2^g \end{matrix}\right\| = H \left\|\begin{matrix} I_1^c \\ I_1^g \\ E_2^c \\ E_2^g \end{matrix}\right\| \qquad (44)$$

$$\left\|\begin{matrix} I_1^c \\ I_1^g \\ E_2^c \\ E_2^g \end{matrix}\right\| = G \left\|\begin{matrix} E_1^c \\ E_1^g \\ I_2^c \\ I_2^g \end{matrix}\right\| \qquad (45)$$

and

$$\left\|\begin{matrix} E_1^c \\ E_1^g \\ I_1^c \\ I_1^g \end{matrix}\right\| = \left\|\begin{matrix} A & -B \\ C & -D \end{matrix}\right\| \left\|\begin{matrix} E_2^c \\ E_2^g \\ I_2^c \\ I_2^g \end{matrix}\right\| \qquad (46)$$

The analytic expressions for the matrix components of equations 43 through 46 can readily be obtained from equation 38 and are given in the Appendix.

The fundamental relations of the basic element have been derived in matrix form, this form being particularly useful in the analysis of magnetic amplifiers resulting from interconnections of basic elements. Two ordinary 2-port networks can be interconnected in five fundamental ways:

1. cascade
2. parallel
3. series
4. series-parallel
5. parallel-series

The analysis of these interconnections reduces to simple matrix addition or multiplication if the proper analytic representation for the individual networks is used.[8] The matrices *Y*, *Z*, *H*, *G*, and *ABCD* were all derived in order to have available the appropriate analytic representation for the various fundamental interconnections. It is pointed out that there is an important restriction to the matrix method, when interconnections *b*, *c*, *d*, and *e* are of interest. If a potential difference exists between the terminals to be joined in the interconnection of two networks, a circulating current will result and the combined behavior will not be given by superposition of the previous individual behaviors. For the basic amplifier under consideration, however, this restriction does not apply, since complete isolation exists between the two ports; and superposition of the behavior of the individual elements can always be used to determine the combined behavior. In the following section, the rules of interconnection of basic elements are developed.

Interconnection of Two Basic Elements

In the previous section it was stated that two 2-port networks can be interconnected in five fundamental ways.

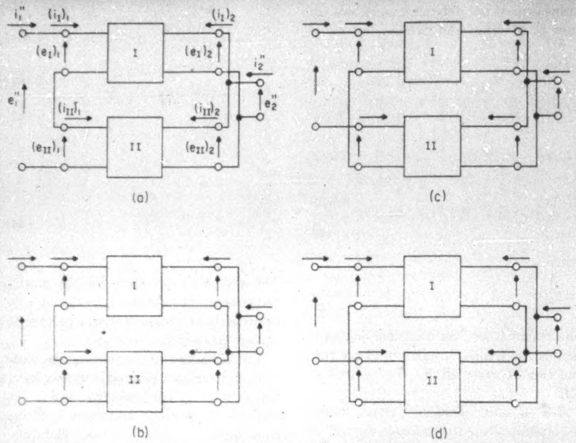


Fig. 4. The four possible subconnections for the series parallel interconnection of 2-port networks

Each of the fundamental interconnections, in turn, can be made in four possible ways, shown in Fig. 4 for the fundamental series-parallel interconnection. As demonstrated by Guillemin,[3] the analytic representation of the interconnected network can easily be obtained by choosing the proper matrix representations of the individual networks. The various subconnections of Fig. 4 can be taken into account by a simple transformation of the variables of one of the networks as will be shown next.

The proper analytic representation for the series-parallel interconnection is given by the H-matrix. Let H_I be the matrix of network I, and H_{II} be the matrix of network II when the port voltage and current polarities are as shown in Fig. 4. The following equations are then valid

For network I,

$$\begin{Vmatrix}(\bar{E}_I)_1\\(\bar{I}_I)_2\end{Vmatrix}=H_I\begin{Vmatrix}(\bar{I}_I)_1\\(\bar{E}_I)_2\end{Vmatrix} \tag{47}$$

For network II,

$$\begin{Vmatrix}(\bar{E}_{II})_1\\(\bar{I}_{II})_2\end{Vmatrix}=H_{II}\begin{Vmatrix}(\bar{I}_{II})_1\\(\bar{E}_{II})_2\end{Vmatrix} \tag{48}$$

where the subscript I refers to network I and the subscript II refers to network II. If the two networks under consideration represent ordinary linear networks, the variable quantities in equations 47 and 48 are the Laplace transform of the instantaneous port voltages and currents. If, however, the two networks represent basic magnetic-amplifier elements, then the varible quantities in equations 47 and 48 are 1×2 submatrices in which the elements are the Laplace transforms of half-period average voltages and currents as defined in the previous section.

The constraints imposed upon the port voltages and currents of the individual networks of the various subconnections are

$$(\bar{E}_I)_2=\pm(\bar{E}_{II})_2 \tag{49}$$

$$(\bar{I}_I)_1=\pm(\bar{I}_{II})_1 \tag{50}$$

where the four possible combinations of sign represent the four possible subconnections.

The port voltages and current of the combination network for the various subconnections can be written in terms of the port voltages and currents of the individual networks as follows:

$$\bar{E}_1''=(\bar{E}_I)_1\pm(\bar{E}_{II})_1 \tag{51}$$

$$\bar{I}_1''=(\bar{I}_I)_1 \tag{52}$$

$$\bar{E}_2''=(\bar{E}_I)_2 \tag{53}$$

$$\bar{I}_2''=(\bar{I}_I)\pm(\bar{I}_{II})_2 \tag{54}$$

where the double-primed variable quantities are the port voltages and currents of the combination network. Note that for a particular subconnection the last terms in equations 49 and 54 have the same sign and also that the last two terms in equations 50 and 51 have the same sign. The sign of the last terms in equations 49 and 54, however, might differ from the sign of the last terms in equations 50 and 51 depending upon the particular subconnection under consideration.

A new set of variables now will be introduced. Let

$$\bar{E}_1'=\pm(\bar{E}_{II})_1 \tag{55}$$

$$\bar{E}_2'=\pm(\bar{E}_{II})_2 \tag{56}$$

and

$$\bar{I}_1'=\pm(\bar{I}_{II})_1 \tag{57}$$

$$\bar{I}_2'=\pm(\bar{I}_{II})_2 \tag{58}$$

where again, for a particular subconnection, the terms in equations 55 and 57 have the same sign, and the terms in equations 56 and 58 have the same sign. Because of this property equations 55 through 58 can be written as follows:

$$\begin{Vmatrix}(\bar{E}_{II})_1\\(\bar{I}_{II})_2\end{Vmatrix}=T\begin{Vmatrix}\bar{E}_1'\\\bar{I}_2'\end{Vmatrix} \tag{59}$$

and

$$\begin{Vmatrix}(\bar{I}_{II})_1\\(\bar{E}_{II})_2\end{Vmatrix}=T\begin{Vmatrix}\bar{I}_1'\\\bar{E}_2'\end{Vmatrix} \tag{60}$$

where for ordinary networks,

$$T=\begin{Vmatrix}\pm1 & 0\\0 & \pm1\end{Vmatrix} \tag{61}$$

This matrix is called the transformation matrix, and the four possible combinations of sign in this matrix represent the four possible subconnections.

If equations 59 and 60 are substituted in equation 48, there results

$$T\begin{Vmatrix}\bar{E}_1'\\\bar{I}_2'\end{Vmatrix}=H_{II}T\begin{Vmatrix}\bar{I}_1'\\\bar{E}_2'\end{Vmatrix} \tag{62}$$

The transformation matrix T is equal to its inverse so that premultiplying equation 62 by the matrix T yields

$$\begin{Vmatrix}\bar{E}_1'\\\bar{I}_2'\end{Vmatrix}=TH_{II}T\begin{Vmatrix}\bar{I}_1'\\\bar{E}_2'\end{Vmatrix} \tag{63}$$

From equations 49, 50, 56, and 57

$$\bar{I}_1'=(\bar{I}_I)_1 \tag{64}$$

$$\bar{E}_2'=(\bar{E}_I)_2 \tag{65}$$

Substitution of equations 55, 58, 64, and 65 in equation 63 yields

$$\begin{Vmatrix}\pm(\bar{E}_{II})_1\\\pm(\bar{I}_{II})_2\end{Vmatrix}=TH_{II}T\begin{Vmatrix}(\bar{I}_I)_1\\(\bar{E}_I)_2\end{Vmatrix} \tag{66}$$

The matrix equation for the subconnections is obtained by adding equations 47 and 66 and substituting equations 51 through 54 into the resulting equation. The following result is obtained

$$\begin{Vmatrix}\bar{E}_1''\\\bar{I}_2''\end{Vmatrix}=(H_I+TH_{II}T)\begin{Vmatrix}\bar{I}_1''\\\bar{E}_2''\end{Vmatrix} \tag{67}$$

Thus, the matrix of the combination network is given by the expression

$$H=H_I+TH_{II}T \tag{68}$$

It can be shown that a similar matrix can be obtained for any one of the other

Vol.
Pt. 1
1959

basic interconnections if the proper matrix relation is used. For example, for the series-series interconnection the impedance matrix becomes

$$Z=Z_I+TZ_{II}T \tag{69}$$

It has already been pointed out that if the networks *I* and *II* represent basic magnetic-amplifier elements, the variable quantities in the foregoing equations are 1×2 submatrices. For magnetic amplifier elements, therefore, the elements in the transformation matrix T (see equation 61) are 2×2 submatrices. For this reason the transformation matrix is written in the form

$$T=\left\|\begin{matrix}\pm X & 0\\ 0 & \pm X\end{matrix}\right\| \tag{70}$$

where the zeros are 2×2 zero submatrices. From the foregoing analysis it may be expected that the matrix X is a 2×2 unit matrix. However, it will be shown that the matrix X also can take the form

$$X=\left\|\begin{matrix}0 & 1\\ 1 & 0\end{matrix}\right\| \tag{71}$$

Practical magnetic amplifiers invariably consist of interconnections of two or more identical basic elements. For two ordinary, linear, identical networks the four possible subconnections reduce to two—subconnections *a* and *b*, for example, in Fig. 4—the other subconnections being identical save for reference directions of the final-network port voltages and currents. For basic magnetic-amplifier elements, however, the four possible subconnections, in general, result in four different magnetic-amplifier circuits even though the two elements are identical, for the following reason. If a d-c signal is applied to the input terminals of basic element *I* in Fig. 5, the saturable inductor of this element saturates during the positive half periods of the supply voltage wave. If the polarity of the input d-c signal is reversed for basic element *II* in Fig. 5, then the saturable inductor of this element saturates during the negative half periods of the supply voltage wave. Thus, if the supply voltage is common to both elements, the inductors of elements *I* and *II* saturate during alternate half periods. For interconnections *c* and *d* in Fig. 4 such a polarity reversal of the input d-c signal is obtained because of the reversal of the input terminals of element *II*. For interconnections *a* and *b*, however, no such reversal exists. Thus, if the supply voltage is common to both elements of the interconnection, the saturable inductors of interconnections *a* and *b* saturate during the same half periods of the supply voltage wave, while the saturable inductors of interconnections *c* and *d* saturate during alternate half periods of the supply voltage wave. This is the property which results in four different magnetic-amplifier circuits for each of the basic interconnections.

The constraints imposed upon the port voltage and currents of the basic elements of interconnections *a* and *b* are (see equations 49 and 50)

$$\left\|\begin{matrix}(E_I)_2^c\\ (E_I)_2^g\end{matrix}\right\|=\pm\left\|\begin{matrix}(E_{II})_2^c\\ (E_{II})_2^g\end{matrix}\right\| \tag{72}$$

and

$$\left\|\begin{matrix}(I_I)_1^c\\ (I_I)_1^g\end{matrix}\right\|=\pm\left\|\begin{matrix}(I_{II})_1^c\\ (I_{II})_1^g\end{matrix}\right\| \tag{73}$$

For subconnections *c* and *d*, however, the control period of element *I* corresponds to the gating period of element *II* and vice versa, so that the constraint equations become

$$\left\|\begin{matrix}(E_I)_2^c\\ (E_I)_2^g\end{matrix}\right\|=\pm\left\|\begin{matrix}(E_{II})_2^g\\ (E_{II})_2^c\end{matrix}\right\| \tag{74}$$

and

$$\left\|\begin{matrix}(I_I)_1^c\\ (I_I)_1^g\end{matrix}\right\|=\pm\left\|\begin{matrix}(I_{II})_1^g\\ (I_{II})_1^c\end{matrix}\right\| \tag{75}$$

Equations 74 and 75 can be rewritten

$$\left\|\begin{matrix}(E_I)_2^c\\ (E_I)_2^g\end{matrix}\right\|=\pm\left\|\begin{matrix}0 & 1\\ 1 & 0\end{matrix}\right\|\left\|\begin{matrix}(E_{II})_2^c\\ (E_{II})_2^g\end{matrix}\right\| \tag{76}$$

and

$$\left\|\begin{matrix}(I_I)_1^c\\ (I_I)_1^g\end{matrix}\right\|=\pm\left\|\begin{matrix}0 & 1\\ 1 & 0\end{matrix}\right\|\left\|\begin{matrix}(I_{II})_1^c\\ (I_{II})_1^g\end{matrix}\right\| \tag{77}$$

If a matrix X is defined so that

$$X=U=\left\|\begin{matrix}1 & 0\\ 0 & 1\end{matrix}\right\| \tag{78}$$

when the two basic elements inductors saturate during the same half periods, and

$$X=\left\|\begin{matrix}0 & 1\\ 1 & 0\end{matrix}\right\| \tag{79}$$

when the two basic elements inductors saturate during alternate half periods, then the constraints imposed upon the port voltages and currents of the basic elements of the subconnections can be expressed in general as follows:

$$(E_I)_2=\pm X(E_{II})_2 \tag{80}$$

$$(I_I)_1=\pm X(I_{II})_1 \tag{81}$$

where the port voltages and currents are 1×2 column matrices representing the division of the operating period into the control and gating periods. The same procedure as used in the derivation of equation 68 can now be followed. The result is

$$H=H_I+TH_IT \tag{82}$$

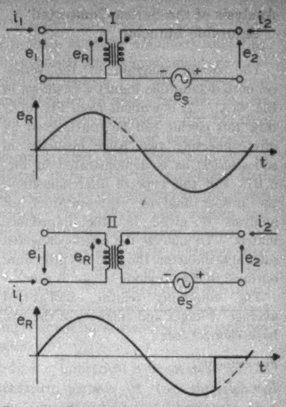


Fig. 5. Saturation during alternate half periods of supply voltage wave

where H is the matrix of the interconnection, H_I is the matrix of the basic element, and T is the transformation matrix

$$T=\left\|\begin{matrix}\pm X & 0\\ 0 & \pm X\end{matrix}\right\| \tag{83}$$

The interchange of control and gating periods considered in the foregoing analysis is caused by the polarity reversal of the applied d-c signal. This polarity reversal is by no means the only factor that causes interchange of control and gating periods. Direct signals at the output port or a 180-degree change in phase of the supply voltage might also cause interchange of control and gating periods. When a particular interconnection is to be studied, all these factors must be taken into account to determine if interchange of control and gating periods occurs.

Interconnections of basic elements that result in interchange of control and gating periods impose an additional restriction upon the voltage waveforms. The matrix addition of equation 82 is valid only if the time periods of elements *I* and *II* in which the average-value quantities are defined coincide; that is, if the control period of element *I* corresponds to the gating period of element *II*, and vice versa. To satisfy this condition it is necessary first to divide the operating period into two equal parts, and second to ensure that the inductor of one element ceases to saturate at the end of the control period of the other element. Clearly, if the waveforms are symmetric, these conditions can be satisfied.

To illustrate the application of the above analysis procedure, the series-connected saturable inductor is analyzed in the next section.

Analysis of the Series-Connected Saturable Inductor

The circuit diagram of the series-connected saturable inductor is shown in Fig. 6(A). It is evident from Fig. 6(A) that this circuit can be constructed by interconnecting two basic elements in series-series, as is demonstrated in Fig. 6(B). For this type of interconnection, the proper analytic representation for analysis is given by the impedance matrix. To analyze the series-connected saturable inductor, the analysis procedure outlined in the previous section is used.

The equations relating half-period average voltages and currents of the two basic elements are:

For element *I*,

$$\bar{E}_I = Z_I \bar{I}_I \tag{84}$$

For element *II*,

$$\bar{E}_{II} = Z_{II} \bar{I}_{II} \tag{85}$$

where the matrices $\bar{E}_I$, $\bar{I}_I$, $\bar{E}_{II}$, and $\bar{I}_{II}$ are 1×4 column matrices of the form

$$\bar{E}_I = \begin{Vmatrix} (\bar{E}_I)_1^c \\ (\bar{E}_I)_1^g \\ (\bar{E}_I)_2^c \\ (\bar{E}_I)_2^g \end{Vmatrix} \tag{86}$$

The two elements are identical so that

$$Z_I = Z_{II} \tag{87}$$

The impedance matrix of the basic element is

$$Z_I = \begin{Vmatrix} R_1+R_m & 0 & R_m & 0 \\ -R_m/z & R_1 & -R_m/z & 0 \\ R_m & 0 & R_2+R_m & 0 \\ -R_m/z & 0 & -R_m/z & R \end{Vmatrix} \tag{88}$$

The next step in the analysis is to determine the transformation matrix *T*. As demonstrated in the previous section, this transformation matrix is uniquely determined by the constraints imposed upon the port voltages and currents of the two elements by the interconnection. Because the input-voltage polarity of element *II* is reversed, the inductors of the two elements saturate during alternate half periods (the subconnection of Fig. 6(B) corresponds to subconnection *c* in Fig. 4 for the series-parallel interconnection). Thus, the control period of element *I* coincides with the gating period of element *II*, and vice versa. The constraint equation becomes

$$\bar{I}_{II} = \begin{Vmatrix} -X & 0 \\ 0 & X \end{Vmatrix} I_I = T I_I \tag{89}$$

where *X* is the submatrix

$$X = \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix} \tag{90}$$

and *T* the desired transformation matrix.

The following transformation of variables is now made:

$$\bar{E}_{II} = T\bar{E}' \tag{91}$$

$$I_{II} = T\bar{I}' \tag{92}$$

Combining equations 85, 87, 91, and 92 yields

$$T\bar{E}' = Z_I T\bar{I}' \tag{93}$$

The matrix *T* is equal to its inverse so that premultiplying equation 93 by *T* yields

$$\bar{E} = T Z_I T \bar{I}' \tag{94}$$

From equations 89 and 92

$$\bar{I}' = I_I \tag{95}$$

Substitution of equation 95 into equation 94 gives

$$\bar{E}' = T Z_I T \bar{I}_I \tag{96}$$

Addition of equations 84 and 96 gives

$$\bar{E}_I + \bar{E}' = (Z_I + T Z_I T)\bar{I}_I \tag{97}$$

But

$$\bar{E}_I + \bar{E}' = \bar{E}'' \tag{98}$$

$$I_I = \bar{I}'' \tag{99}$$

where the doubled primed quantities are the port voltages and current of the combination network. Thus, substitution of equations 98 and 99 into equation 97 yields the following equation:

$$\bar{E}'' = (Z_I + T Z_I T)\bar{I}'' = Z\bar{I}'' \tag{100}$$

Substitution of equation 88 in equation 100 and performing the matrix operation indicated by equation 100 gives

$$Z = \begin{Vmatrix} 2R_1+R_m & -R_m/z & R_m & R_m/z \\ -R_m/z & 2R_1+R_m & -R_m/z & -R_m \\ R_m & R_m/z & 2R_2+R_m & -R_m/z \\ -R_m/z & -R_m & -R_m/z & 2R_2+R_m \end{Vmatrix} \tag{101}$$

A very interesting property is revealed by equation 101 if the matrix is partitioned into four 2×2 submatrices. The elements along the two principal diagonals of each of these submatrices are seen to be equal in magnitude. If the transforma- is made,

$$\bar{E}'' = A\bar{E}''' \tag{102}$$

$$\bar{I}'' = A\bar{I}''' \tag{103}$$

where

$$A = \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix} \tag{104}$$

the matrix equation, equation 105, is obtained.

$$\begin{Vmatrix} (\bar{E}''')_1^c \\ (\bar{E}''')_1^g \\ (\bar{E}''')_2^c \\ (\bar{E}''')_1^g \end{Vmatrix} = \begin{Vmatrix} 2R_1+R_m & -R_m/z & -R_m & R_m/z \\ -R_m/z & 2R_1+R_m & R_m/z & -R_m \\ -R_m & -R_m/z & 2R_2+R_m & R_m/z \\ -R_m/z & -R_m & R_m/z & 2R_2+R_m \end{Vmatrix} \begin{Vmatrix} (\bar{I}''')_1^c \\ (\bar{I}''')_1^g \\ (\bar{I}''')_2^c \\ (\bar{I}''')_2^g \end{Vmatrix} \tag{105}$$

Addition of the first two and last two rows in equation 105 yields

$$\bar{E}_1 = [2R_1 + R_m(1-1/z)]\bar{I}_1 + R_m(-1+1/z)\bar{I}_2 \tag{106}$$

$$\bar{E}_2 = -R_m(1+1/z)\bar{I}_1 + [2R_2 + R_m(1+1/z)]\bar{I}_2 \tag{107}$$

where

$$\bar{E}_1 = (\bar{E}''')_1^c + (\bar{E}''')_1^g$$

$$\bar{E}_2 = (\bar{E}''')_2^c + (\bar{E}''')_2^g$$

$$\bar{I}_1 = (\bar{I}''')_1^c + (\bar{I}''')_1^g$$

$$\bar{I}_2 = (\bar{I}''')_2^c + (\bar{I}''')_2^g \tag{108}$$

The new variables introduced by equation 108 can be interpreted as follows. The independent current variables with superscript *c* are by definition zero during the gating periods. Similarly, the independent current variables with superscript *g* are by definition zero during the control periods. It is evident from equation 105 that the dependent voltage variables also possess the same property because the terms in the expression for a dependent-voltage variable involving currents of superscripts different from that of the voltage variable are shifted in time one half period. Thus, during the control period

$$\bar{E}_1 = (\bar{E}''')_1^c$$

$$\bar{E}_2 = (\bar{E}''')_2^c$$

$$\bar{I}_1 = (\bar{I}''')_1^c$$

$$\bar{I}_2 = (\bar{I}''')_2^c \tag{109}$$

and during the gating period

$$\bar{E}_1 = (\bar{E}''')_1^g$$

$$\bar{E}_2 = (\bar{E}''')_2^g$$

$$\bar{I}_1 = (\bar{I}''')_1^g$$

$$\bar{I}_2 = (\bar{I}''')_2^g \tag{110}$$

The new variables introduced by equation 108, therefore, are half-period average voltages and currents. These variables are related to the half-period average voltages and currents of the series-connected saturable inductor by a transformation of the type expressed by equations 102 and 103.

The transformation expressed by equations 102 and 103 is fundamental to the analysis of magnetic amplifiers, particularly magnetic amplifiers with diodes. Physically, the transformation can be interpreted as representing the action of a synchronous switch, the switch being operated so that if a d-c signal is applied to it, an a-c output signal is obtained. Or, if an a-c signal is applied, a d-c output signal is obtained. Equation 105, therefore, represents the series-connected saturable inductor in cascade with such a synchronous switch.

If the series-connected saturable inductor is terminated in a linear, passive load, the steady-state output consists of an a-c signal when a d-c signal is applied to the input. This property can be seen from the circuit diagram of Fig. 6. In steady-state operation the voltage across the saturable inductor terminals cannot contain a direct-voltage component. Thus, the voltage across the series combination of the load and the resistance R_2 must be an alternating voltage. The transformation of equations 102 and 103, therefore, can be represented physically as the action of a full-wave rectifier connected to the output terminals. The full wave rectifier renders the series-connected saturable inductor symmetric so that no distinction can be made between the different half periods. The four amplifier equations, therefore, can be reduced to two.

To derive the transfer characteristic of the series-connected saturable inductor with resistive load, the following additional equation is used:

$$E_2 = -R_L I_2 \qquad (111)$$

After equation 111 is substituted into equations 106 and 107 and the resulting equations are solved for E_2 in terms of E_1, equation 112 is obtained by letting R_m tend to infinity and R_2 tend to zero:

$$zE_2 = -\frac{R_L}{R_L+2R_1}(z+1)E_1 + \frac{R_L-2R_1}{R_L+2R_1}E_2 \qquad (112)$$

The inverse Laplace transform of equation 112 produces the same difference equation for the series-connected saturable-inductor circuit as was previously derived by the author.[4]

$$E_2(n+1) = -\frac{R_L}{R_L+2R_1}[E_1(n+1)+E_1(n)] + \frac{R_L-2R_1}{R_L+2R_1}E_2(n) \qquad (113)$$

Since for this type of circuit the magnetizing resistance, R_m, is very large compared with resistance, R_1, and the resistance, R_2, is small compared with load resistance, R_L, the limit process performed in the derivation of equation 112 is applicable to practical circuits.

Interconnection of *n* Basic Elements

The analysis developed previously for a single-phase magnetic amplifier can readily be extended to include the polyphase case. Characteristic of the single-phase amplifier analysis are the division of the operating period into the control and gating periods, and the selection of half-period average voltages and currents as variables. The characteristic equations written then relate variables in the different half periods of operation. Equations written for individual basic elements may be combined in a simple fashion to obtain equations for an interconnection of basic elements, provided that the control and gating periods of the individual elements have the same position in time, or occur so that the control period of one element corresponds to the gating period of the other.

An *n*-phase magnetic amplifier consists of *n* single-phase magnetic amplifiers each connected to one phase of a symmetrical *n*-phase power source. The single-phase amplifiers are usually alike and may consist of one or more cores. The analysis to be presented, however, is not restricted to *n*-phase magnetic amplifiers consisting of *n* identical single-phase amplifiers but is valid in general. If the single-phase amplifiers are all alike, then the analysis can be greatly simplified. For the case of identical single-phase amplifiers in a power system with even numbers of phases, there exists the possibility that the outputs of the individual amplifiers cancel. The analysis still applies and shows that this cancellation occurs. In regard to the steady-state characteristics, it is of interest to note that, if the quiescent currents cancel in the individual single-phase amplifiers, then the *n*-phase amplifier has linear steady-state characteristics.

In a symmetrical *n*-phase system, the operating periods of two single-phase amplifiers in subsequent phases are displaced in time from each other by T/n. The control and gating periods of the individual single-phase amplifiers, therefore, do not have the same position in time nor are they shifted an integral number of half-periods with respect to each other. To overcome this difficulty, it is necessary to divide the control and gating periods into subintervals. In the analysis that follows, it is assumed for an *n*-phase amplifier that each of these periods is divided into *n* subintervals. This subdivision is illustrated in Fig. 7. It may be seen from the figure that the operating periods of the individual single-phase amplifiers are shifted an integral number of subintervals with respect to each other; thus, time coincidence of the subintervals of the various single-phase amplifiers is obtained. The shift between successive phases in Fig. 7 is two subintervals. Therefore, if *n* is even, a possible alternative is to divide the control and gating periods into only *n*/2 subintervals. The larger number, *n*, of subintervals is used so that the

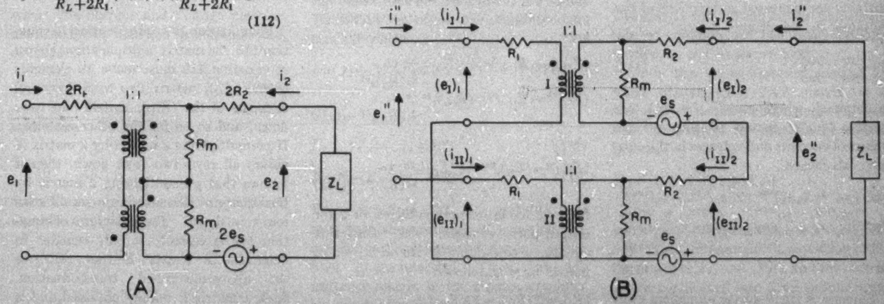


Fig. 6. Series-connected saturable inductor. B—As an interconnection of two basic elements

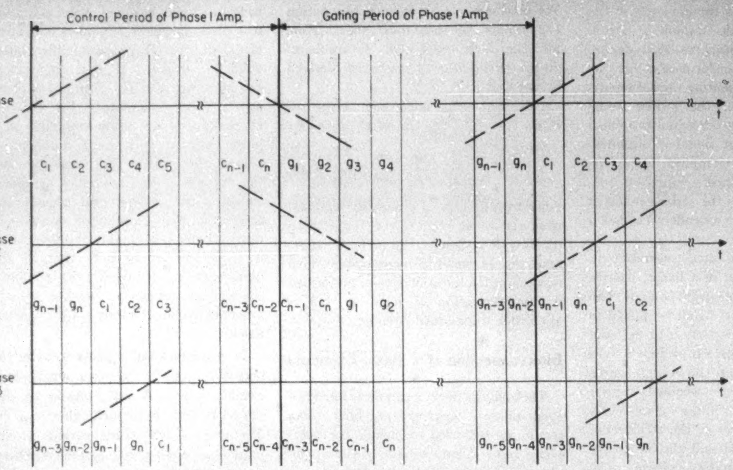

Control Period of Phase I Amp.
Gating Period of Phase I Amp.
Phase 1
c_1 c_2 c_3 c_4 c_5 c_{n-1} c_n g_1 g_2 g_3 g_4 g_{n-1} g_n c_1 c_2 c_3 c_4
Phase 2
g_{n-1} g_n c_1 c_2 c_3 c_{n-3} c_{n-2} c_{n-1} c_n g_1 g_2 g_{n-3} g_{n-2} g_{n-1} g_n c_1 c_2
Phase 3
g_{n-3} g_{n-2} g_{n-1} g_n c_1 c_{n-5} c_{n-4} c_{n-3} c_{n-2} c_{n-1} c_n g_{n-5} g_{n-4} g_{n-3} g_{n-2} g_{n-1} g_n
t
t
t

Fig. 7. Division into n subintervals of the control and gating periods of the amplifiers in their successive phases of an n-phase magnetic amplifier

analysis may be applicable whether n is even or odd.

To analyze the n-phase magnetic amplifier, each of the single-phase magnetic amplifiers is described by a set of equations relating average voltages and currents in the various subintervals, the number of equations being four times (two for control and gating subintervals two for input and output ports) the number of subintervals n. For example, in a symmetrical 3-phase system the number of subintervals is three; thus, twelve equations are needed to describe the operation of the individual single-phase amplifiers. The $4n$ equations may be written in matrix form. The matrix of the equations may take any of the forms previously introduced (Y, Z, G, H, or $ABCD$), depending upon which currents and voltages are expressed as functions of the other variables. If, for example, the currents are expressed as functions of the voltages, the admittance matrix

$$Y=\left\|\begin{matrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{matrix}\right\| \qquad (114)$$

is obtained, where each y_{rs} $(r,s=1,2)$ is itself a $2n\times 2n$ matrix relating average values of currents and voltages in the subintervals chosen.

$$y_{rs}=\left\|\begin{matrix} y_{rs}^{c_1c_1}\ldots y_{rs}^{c_1c_n} & y_{rs}^{c_1g_1}\ldots y_{rs}^{c_1g_n} \\ y_{rs}^{c_nc_1}\ldots y_{rs}^{c_nc_n} & y_{rs}^{c_ng_1}\ldots y_{rs}^{c_ng_n} \\ y_{rs}^{g_1c_1}\ldots y_{rs}^{g_1c_n} & y_{rs}^{g_1g_1}\ldots y_{rs}^{g_1g_n} \\ y_{rs}^{g_nc_1}\ldots y_{rs}^{g_nc_n} & y_{rs}^{g_ng_1}\ldots y_{rs}^{g_ng_n} \end{matrix}\right\| \qquad (115)$$

Suppose n single-phase amplifiers each connected to one phase of an n-phase power source are interconnected parallel-parallel. If Y_I, Y_{II}, … Y_N are the admittance matrices for the individual amplifiers, then the equations for these amplifiers become

$$\begin{aligned} I_I &= Y_I\bar{E}_I \\ I_{II} &= Y_{II}\bar{E}_{II} \\ &\cdot\quad\cdot\quad\cdot \\ I_N &= Y_N\bar{E}_N \end{aligned} \qquad (116)$$

where the current and voltage matrices are $1\times 4n$ column matrices.

As seen in Fig. 7, the following subintervals of operation correspond in time:

$$\begin{array}{lllllll} \text{Ampl. } I: & c_1 & c_2 & c_3 & \ldots c_n & g_1 & \ldots g_n \\ \text{Ampl. } II: & g_{n-1} & g_n & c_1 & \ldots c_{n-2} & c_{n-1} & \ldots g_{n-2} \\ \text{Ampl. } III: & g_{n-3} & g_{n-2} & g_{n-1} & \ldots c_{n-4} & c_{n-3} & \ldots g_{n-4} \\ & \cdot & \cdot & \cdot & \cdots\cdot & \cdot & \cdots\cdot \\ \text{Ampl. } N: & c_3 & c_4 & c_5 & \ldots . g_2 & g_3 & \ldots . c_2 \end{array} \qquad (117)$$

Thus, the constraints imposed upon the port voltages by the interconnection are:

$$\begin{aligned} (\bar{E}_I)_s^{c_1} &= (\bar{E}_{II})_s^{g_{n-1}} = (\bar{E}_{III})_s^{g_{n-3}} = \ldots = (\bar{E}_N)_s^{c_3} \\ (\bar{E}_I)_s^{c_2} &= (\bar{E}_{II})_s^{g_n} = (\bar{E}_{III})_s^{g_{n-2}} = \ldots = (\bar{E}_D)_s^{c_4} \\ &\cdot\qquad\cdot\qquad\cdot\qquad\cdots \\ (\bar{E}_I)_s^{g_n} &= (\bar{E}_{II})_s^{g_{n-2}} = (\bar{E}_{III})_s^{g_{n-4}} = \ldots = (\bar{E}_N)_s^{c_2} \end{aligned} \qquad (118)$$

for $s=1$ for input-port voltages, or 2 for output-port voltages. The constraint equations, equation 118, can be written in matrix form as follows:

$$(\bar{E}_I)_s = X(\bar{E}_{II})_s = X^2(\bar{E}_{III})_s = \ldots = X^{n-1}(\bar{E}_N)_s \qquad (119)$$

where the voltage matrices are $1\times 2n$ column matrices of the form

$$(\bar{E}_I)_s=\left\|\begin{matrix} (\bar{E}_I)_s^{c_1} \\ (\bar{E}_I)_s^{c_2} \\ \cdot \\ \cdot \\ \cdot \\ (\bar{E}I)_s^{c_n} \\ (\bar{E}_I)_s^{g_1} \\ \cdot \\ \cdot \\ \cdot \\ (\bar{E}_I)_s^{g_n} \end{matrix}\right\| \qquad (120)$$

For equation 119 to be identical to equation 118, the matrix multiplications shown in equation 119 must move all elements of the $(\bar{E}_{II})_s$ matrix two rows down, all elements of the $(\bar{E}_{III})_s$ matrix four rows down, and so on for the other matrices. If premultiplying a matrix by a matrix X moves all rows two rows down, then it follows that premultiplying a matrix by the square of this matrix moves all rows four rows down. Thus, the form of equation 119 is correct; it only remains to determine a matrix X that performs the above-mentioned transformation. Such a matrix is readily obtained and is as follows:

Vol.
Pt. 1
1959

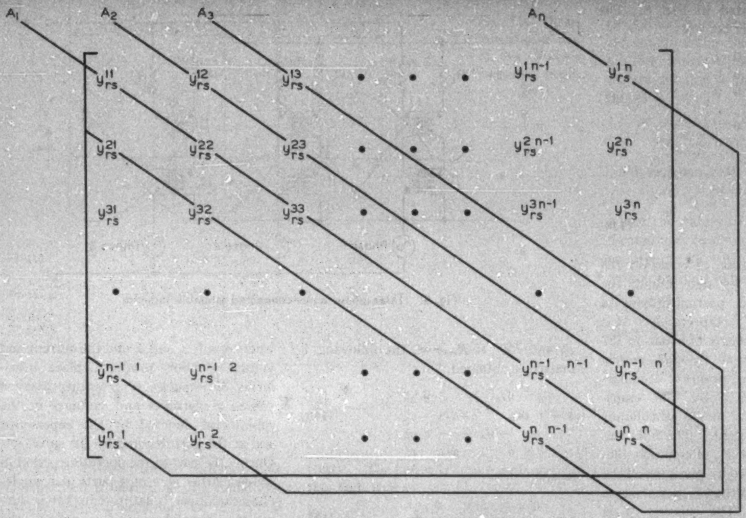


Fig. 8. Addition of matrix components. The elements in the matrix at the left are 2×2 submatrices of the form:

$$y_{rs}=\left\Vert\begin{matrix} y_{rs}^{c1c3} & y_{rs}^{c1c4} \\ y_{rs}^{c2c3} & y_{rs}^{c2c4}\end{matrix}\right\Vert$$

$$X=\left\Vert\begin{matrix} 0 & 0 & . & . & . & & & 1 & 0 \\ 0 & 0 & . & . & . & & & 0 & 1 \\ 1 & 0 & . & . & . & & & 0 & 0 \\ 0 & 1 & 0 & . & . & & & 0 & 0 \\ 0 & 0 & 1 & 0 & . & & & 0 & 0 \\ . & . & . & . & . & & & . & . \\ 0 & 0 & . & . & . & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & . & . & . & & 0 & 1 & 0 & 0 \end{matrix}\right\Vert \qquad (121)$$

In order to derive the matrix relation for the interconnection the following new variables are introduced.

$$\begin{aligned} E_{II}' &= TE_{II} \\ E_{III}' &= T^2E_{III} \\ &\cdot \quad \cdot \quad \cdot \\ E_N' &= T^{n-1}E_N \end{aligned} \qquad (122)$$

and

$$\begin{aligned} I_{II}' &= TI_{II} \\ &\cdot \quad \cdot \quad \cdot \\ I_N' &= T^{n-1}I_N \end{aligned} \qquad (123)$$

where the current and voltage matrices are the 1×4*n* matrices in equation 116, of the form

$$E_{II}=\left\Vert\begin{matrix}(E_{II})_1 \\ (E_{II})_2\end{matrix}\right\Vert \qquad (124)$$

and where

$$T=\left\Vert\begin{matrix}X & 0 \\ 0 & X\end{matrix}\right\Vert \qquad (125)$$

Substitution of equations 122 and 123 into equation 116 yields

$$\begin{aligned} I_I &= Y_IE_I \\ I_{II} &= TY_{II}T_TE_{II}' \\ &\cdot \quad \cdot \quad \cdot \\ I_N' &= T^{n-1}Y_NT_T^{n-1}E_N' \end{aligned} \qquad (126)$$

In the derivation of equation 126 the identity

$$T^{-1}=T_T \qquad (127)$$

is used. The subscript *T* denotes the transpose of a matrix. The validity of equation 127 can be seen from equations 121 and 125, because

$$T_T=\left\Vert\begin{matrix}X & 0 \\ 0 & X\end{matrix}\right\Vert_T=\left\Vert\begin{matrix}X_T & 0 \\ 0 & X_T\end{matrix}\right\Vert \qquad (128)$$

From equation 121

$$X_T=X^{-1} \qquad (129)$$

so that

$$T_T=\left\Vert\begin{matrix}X^{-1} & 0 \\ 0 & X^{-1}\end{matrix}\right\Vert=T^{-1} \qquad (130)$$

It is evident from equations 119, 122, and 125 that the voltage matrices on the right-hand side of equation 126 are equal. Thus, addition of the equations in equation 126 yields

$$I=(Y_I+TY_{II}T_T+\ldots+T^{n-1}Y_NT_T^{n-1})E \qquad (131)$$

where

$$I=I_I+I_{II}'+\ldots+I_N' \qquad (132)$$

$$E=E_I=E_{II}'=\ldots=E_N' \qquad (133)$$

From equations 123, 132, and 133 it is seen that the variables of equation 131 are the port voltages and currents of the final interconnection. Thus, the admittance matrix of the *n*-phase amplifier is given by the expression

$$Y=Y_I+TY_{II}T_T+\ldots+T^{n-1}Y_NT_T^{n-1} \qquad (134)$$

The expressions for the other basic interconnections of *n* single-phase amplifiers can be obtained through a similar procedure.

Most practical polyphase magnetic amplifiers consist of interconnections of identical single-phase amplifiers. These single-phase amplifiers are usually interconnected so that there is no polarity reversal of port voltages of currents. For such polyphase amplifiers, therefore, the matrices of the individual single-phase amplifiers are identical. Under these conditions, equation 134 for the admittance matrix of the parallel-parallel interconnection becomes

$$Y=Y_I+TY_IT_T+\ldots+T^{n-1}Y_IT_T^{n-1} \qquad (135)$$

The matrix summation on the right-hand side of equation 135 can easily be performed, as will be shown next.

Since the transformation matrix *T*, when written as a 2×2 matrix, is a diagonal one, the submatrices Y_{rs} (see equation 114) of the admittance matrix become

$$\begin{aligned} Y_{rs}=(Y_I)_{rs}+X(Y_I)_{rs}X_T+\ldots+ \\ X^{n-1}(Y_I)_{rs}X_T^{n-1} \end{aligned} \qquad (136)$$

where *r* and *s* may each be 1 or 2. To perform the summation of equation 136, the matrices Y_{rs} and $(Y_I)_{rs}$ and the matrix *X* are partitioned into 2×2 submatrices. The partitioning of the admittance matrices is illustrated in Fig. 8. The partitioned X matrix becomes

$$X=\begin{Vmatrix}0 & 0 & \ldots & & 0 & U\\ U & 0 & \ldots & & 0 & 0\\ 0 & U & \ldots & & 0 & 0\\ \cdot & \cdot & & & \cdot & \cdot\\ 0 & 0 & \ldots & U & 0 & 0\\ 0 & 0 & \ldots & 0 & U & 0\end{Vmatrix} \tag{137}$$

where the zeros are 2×2 matrices and U is

$$U=\begin{Vmatrix}1 & 0\\ 0 & 1\end{Vmatrix} \tag{138}$$

The matrix summation of equation 136 can be performed by determining the effect of pre- and postmultiplying a matrix by X and X_T, respectively. It is evident from the matrix of equation 137 that premultiplication by the matrix X moves all rows one row down. Similarly, postmultiplication by the transpose of the matrix X moves all columns one column to the right. The final elements in the matrix Y_{rs} of equation 136, therefore, are obtained by summing all the diagonal elements of the $(Y_I)_{rs}$-matrix, as shown by Fig. 8. The subscript I is omitted in Fig. 8 to simplify the figure. Elements in the same diagonal of the final matrix are identical, and the desired matrix is

$$Y_{rs}=\begin{Vmatrix}A_1 & A_2 & \ldots & A_n\\ A_n & A_1 & \ldots & A_{n-1}\\ \cdot & \cdot & \cdot\ \cdot & \cdot\\ A_3 & A_4 & \ldots & A_2\\ A_2 & A_3 & \ldots & A_1\end{Vmatrix} \tag{139}$$

where each A_i $(i=1, 2 \ldots n)$ is a 2×2 matrix obtained in the manner indicated by Fig. 8.

The procedure described will be demonstrated in the next section by applying it to a 3-phase amplifier.

Analysis of 3-Phase Amplifier

The amplifier to be analyzed, shown in Fig. 9, consists of three series-connected saturable inductors interconnected in parallel-parallel. For the parallel-parallel interconnection the proper analytic representation for analysis is the admittance matrix. The first step in the analysis is to determine the short-circuit driving-point and transfer admittances for the single-phase series-connected saturable inductor. When these admittances have been determined, the 3-phase matrix representation of the individual amplifiers will be derived.

The admittance matrix for the series-connected saturable inductor can be found by inverting the impedance matrix of equation 101. To simplify the expressions, it will be assumed that $R_m >> R_1$ and R_2. If $R_m \rightarrow \infty$, the following Y matrix is obtained

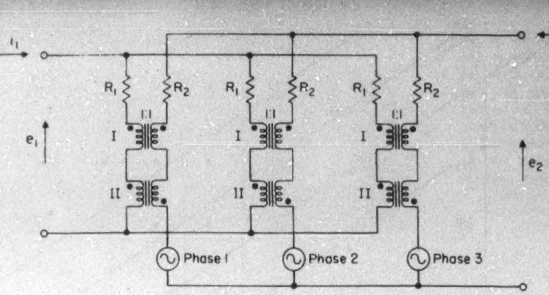


Fig. 9. Three-phase series-connected saturable inductor

$$Y=\begin{Vmatrix}a & b/z & c & d/z\\ b/z & a & -d/z & -c\\ -a & -b/z & -c- & d/z\\ b/z & a & -d/z & -c\end{Vmatrix} \tag{140}$$

where

$$a=\frac{(R_1+R_2)-(R_1-R_2)/z^2}{(R_1+R_2)^2-(R_1-R_2)^2/z^2} \tag{141}$$

$$b=\frac{2R_2}{(R_1+R_2)^2-(R_1-R_2)^2/z^2} \tag{142}$$

$$c=-\frac{(R_1+R_2)+(R_1-R_2)/z^2}{(R_1+R_2)^2-(R_1-R_2)^2/z^2} \tag{143}$$

$$d=-\frac{2R_1}{(R_1+R_2)^2-(R_1-R_2)^2/z^2} \tag{144}$$

To determine the 3-phase matrix representation of the series-connected saturable inductor, it is necessary to establish the subintervals in which saturation of the individual inductors occur. In Fig. 10 these subintervals are shown; in the analysis to be presented the principle mode of operation in which saturation occurs only within the g_3 subinterval of each core in each amplifier is considered. If it is desired to analyze the amplifier when saturation occurs in additional subintervals, e.g., g_2 and g_3 in Fig. 10, a new set of equations must be written.

For the individual amplifiers in the different phases, equations can be written as follows:

$$I_1^{cm}=\sum_{n=1}^{3}\left\Vert y_{11}^{cmcn}E_1^{cn}+y_{11}^{cmgn}E_1^{gn}+y_{12}^{cmcn}E_2^{cn}+y_{12}^{cmgn}E_2^{gn}\right\Vert \tag{145}$$

$$I_1^{gm}=\sum_{n=1}^{3}\left\Vert y_{11}^{gmcn}E_1^{cn}+y_{11}^{gmgn}E_1^{gn}+y_{12}^{gmcn}E_2^{cn}+y_{12}^{gmgn}E_2^{gn}\right\Vert \tag{146}$$

$$I_2^{cm}=\sum_{n=1}^{3}\left\Vert y_{21}^{cmcn}E_1^{cn}+y_{21}^{cmgn}E_1^{gn}+y_{22}^{cmcn}E_2^{cn}+y_{22}^{cmgn}E_2^{gn}\right\Vert \tag{147}$$

$$I_2^{gm}=\sum_{n=1}^{3}\left\Vert y_{21}^{gmcn}E_1^{cn}+y_{21}^{gmgn}E_1^{gn}+y_{22}^{gmcn}E_2^{cn}+y_{22}^{gmgn}E_2^{gn}\right\Vert \tag{148}$$

where $m=1$, 2, and 3, and the current and voltage symbols denote Laplace transforms of impulses representing average values of currents and voltages in the subinterval denoted by the superscript and at the port denoted by the subscript. During the part of the operating period in which neither core of a particular single-phase amplifier is saturated, that is during the intervals c_1, c_2, g_1, and g_2, the driving-point and transfer admittances of that amplifier are zero. Hence, the part currents during these intervals must also be zero.

$$I_1^{c_1}=I_1^{c_2}=I_1^{g_1}=I_1^{g_2}=I_2^{c_1}=I_2^{c_2}=I_2^{gc_1}=I_2^{g_2}=0 \tag{149}$$

To permit the use of the Laplace transform in the analysis of magnetic amplifiers, it was assumed that the variable quantities consist of impulse. It was pointed out that for the purpose of analysis any waveform could be assigned to the variable quantities; the impulse function being chosen for convenience only. For the same reason it is assumed in equations 145 through 149, that the port voltages and currents for the 3-phase amplifier consist of impulses occurring during the subinterval c_1, c_2, c_3 and g_1, g_2, g_3.

The equations for the single-phase

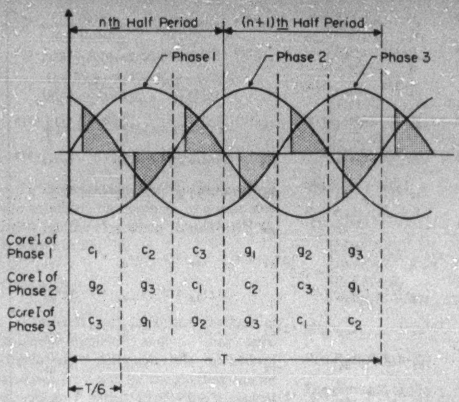


Fig. 10. Intervals of operation of the 3-phase magnetic amplifier

series-connected saturable inductor now can be used to determine the remainder of the port currents for the individual series-connected saturable inductor in the 3-phase system. Suppose that the half-period quantities are represented by impulses that occur during the subinterval c_3 (or g_3), then

$$\bar{I}_1^c=\frac{1}{3}(\bar{I}_1^{c_1}z^{-2/3}+\bar{I}_1^{c_2}z^{-1/3}+\bar{I}_1^{c_3}) \tag{150}$$

where $\bar{I}_1^{c_1}$, $\bar{I}_1^{c_2}$, and $\bar{I}_1^{c_3}$ are the Laplace transform of impulses occurring during the subintervals c_1, c_2, and c_3, respectively. The factor 1/3 takes into account the fact that

$$\bar{I}_1^c$$

is the average current over the complete half period, while

$$\bar{I}_1^{c_1},\ \bar{I}_1^{c_2}, \text{ and } \bar{I}_1^{c_3}$$

are the average currents over the subintervals c_1, c_2, and c_3. Similar relations can also be written for the remainder of the half-period average voltages and currents.

If equation 150 and similar relations for the voltage variables are substituted into the equations for the single-phase amplifier for which equation 140 gives the matrix, the first of the four equations obtained is

$$\begin{aligned}\bar{I}_1^{c_1}z^{-2/3}+\bar{I}_1^{c_2}z^{-1/3}+\bar{I}_1^{c_3}\\ =y_{11}^{cc}(\bar{E}_1^{c_1}z^{-2/3}+\bar{E}_1^{c_2}z^{-1/3}+\bar{E}_1^{c_3})+\\ y_{11}^{cg}(\bar{E}_1^{g_1}z^{-2/3}+\bar{E}_1^{g_2}z^{-1/3}+\bar{E}_1^{g_3})+\\ y_{12}^{cc}(\bar{E}_2^{c_1}z^{-2/3}+\bar{E}_2^{c_2}z^{-1/3}+\bar{E}_2^{c_3})+\\ y_{12}^{cg}(\bar{E}_2^{g_1}z^{-2/3}+\bar{E}_2^{g_2}z^{-1/3}+\bar{E}_2^{g_3})\end{aligned} \tag{151}$$

where the values of the admittances indicated may be determined by comparing equation 140 with equation 40. But from equation 149 both

$$\bar{I}_1^{c_1} \text{ and } \bar{I}_1^{c_2}$$

are zero, thus

$$\begin{aligned}\bar{I}_1^{c_3}=y_{11}^{cc}z^{-2/3}\bar{E}_1^{c_1}+y_{11}^{cc}z^{-1/3}\bar{E}_1^{c_2}+\\ y_{11}^{cc}\bar{E}_1^{c_3}+y_{11}^{cg}z^{-2/3}\bar{E}_1^{g_1}+\\ y_{11}^{cg}z^{-1/3}\bar{E}_1^{g_2}+y_{11}^{cg}\bar{E}_1^{g_3}+\\ y_{12}^{cc}z^{-2/3}\bar{E}_2^{c_1}+y_{12}^{cc}z^{-1/3}\bar{E}_2^{c_2}+\\ y_{12}^{cc}\bar{E}_2^{c_3}+y_{12}^{cg}z^{-1/3}\bar{E}_2^{g_1}+\\ y_{12}^{cg}z^{-1/3}\bar{E}_2^{g_2}+y_{12}^{cg}\bar{E}_2^{g_3}\end{aligned} \tag{152}$$

The remaining three equations obtained from equation 140 lead to similar expressions for

$$\bar{I}_1^{g_3},\ \bar{I}_2^{c_3}, \text{ and } \bar{I}_2^{g_3}$$

In Fig. 11 the complete Y matrix representing the series-connected saturable inductor in a 3-phase system is shown.

The matrix of the interconnected 3-phase amplifier can now be derived by following the procedure outlined in the previous section; the result is shown in Fig. 12. For the present, ignore the signs in parentheses.

$$\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
az^{-2/3} & az^{-1/3} & a & bz^{-5/3} & bz^{-4/3} & bz^{-1} & cz^{-2/3} & cz^{-1/3} & c & dz^{-5/3} & dz^{-4/3} & dz^{-1}\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
bz^{-5/3} & bz^{-4/3} & bz^{-1} & az^{-2/3} & az^{-1/3} & a & -dz^{-5/3} & -dz^{-4/3} & -dz^{-1} & -cz^{-2/3} & -cz^{-1/3} & -c\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
-az^{-2/3} & -az^{-1/3} & -a & -bz^{-5/3} & -bz^{-4/3} & -bz^{-1} & -cz^{-2/3} & -cz^{-1/3} & -c & -dz^{-5/3} & -dz^{-4/3} & -dz^{-1}\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
bz^{-5/3} & bz^{-4/3} & bz^{-1} & az^{-2/3} & az^{-1/3} & a & -dz^{-5/3} & -dz^{-4/3} & -dz^{-1} & -cz^{-2/3} & -cz^{-1/3} & -c
\end{bmatrix}$$

Fig. 11. Admittance matrix of series-connected saturable inductor in a 3-phase system

$$\begin{bmatrix}
a & bz^{-5/3} & bz^{-4/3} & bz^{-1} & az^{-2/3} & az^{-1/3} & c & (-)dz^{-5/3} & dz^{-4/3} & (-)dz^{-1} & cz^{-2/3} & (-)cz^{-1/3} \\
az^{-1/3} & a & bz^{-5/3} & bz^{-4/3} & bz^{-1} & az^{-2/3} & -cz^{-1/3} & (\pm)c & -dz^{-5/3} & (\pm)dz^{-4/3} & -dz^{-1} & (\pm)cz^{-2/3} \\
az^{-2/3} & az^{-1/3} & a & bz^{-5/3} & bz^{-4/3} & bz^{-1} & cz^{-2/3} & (-)cz^{-1/3} & c & (-)dz^{-5/3} & dz^{-4/3} & (-)dz^{-1} \\
bz^{-1} & az^{-2/3} & az^{-1/3} & a & bz^{-5/3} & bz^{-4/3} & -dz^{-1} & (\pm)cz^{-2/3} & -cz^{-1/3} & (\pm)c & -dz^{-5/3} & (\pm)dz^{-4/3} \\
bz^{-4/3} & bz^{-1} & az^{-2/3} & az^{-1/3} & a & bz^{-5/3} & dz^{-4/3} & (-)dz^{-1} & cz^{-2/3} & (-)cz^{-1/3} & c & (-)dz^{-5/3} \\
bz^{-5/3} & bz^{-4/3} & bz^{-1} & az^{-2/3} & az^{-1/3} & a & -dz^{-5/3} & (\pm)dz^{-4/3} & -dz^{-1} & (\pm)cz^{-2/3} & -cz^{-1/3} & (\pm)c \\
-a & -bz^{-5/3} & -bz^{-4/3} & -bz^{-1} & -az^{-2/3} & -az^{-1/3} & -c & (\pm)dz^{-5/3} & -dz^{-4/3} & (\pm)dz^{-1} & -cz^{-2/3} & (\pm)cz^{-1/3} \\
(-)az^{-1/3} & (-)a & (-)bz^{-5/3} & (-)bz^{-4/3} & (-)bz^{-1} & (-)az^{-2/3} & (\pm)cz^{-1/3} & -c & (\pm)dz^{-5/3} & -dz^{-4/3} & (\pm)dz^{-1} & -cz^{-2/3} \\
-az^{-2/3} & -az^{-1/3} & -a & -bz^{-5/3} & -bz^{-4/3} & -bz^{-1} & -cz^{-2/3} & (\pm)cz^{-1/3} & -c & (\pm)dz^{-5/3} & -dz^{-4/3} & (\pm)dz^{-1} \\
(-)bz^{-1} & (-)az^{-2/3} & (-)az^{-1/3} & (-)a & (-)bz^{-5/3} & (-)bz^{-4/3} & (\pm)dz^{-1} & -cz^{-2/3} & (\pm)cz^{-1/3} & -c & (\pm)dz^{-5/3} & -dz^{-4/3} \\
-bz^{-4/3} & -bz^{-1} & -az^{-2/3} & -az^{-1/3} & -a & -bz^{-5/3} & -dz^{-4/3} & (\pm)dz^{-1} & -cz^{-2/3} & (\pm)cz^{-1/3} & -c & (\pm)dz^{-5/3} \\
(-)bz^{-5/3} & (-)bz^{-4/3} & (-)bz^{-1} & (-)az^{-2/3} & (-)az^{-1/3} & (-)a & (\pm)dz^{-5/3} & -dz^{-4/3} & (\pm)dz^{-1} & -cz^{-2/3} & (\pm)cz^{-1/3} & -c
\end{bmatrix}$$

Fig. 12 (above). Matrix of 3-phase amplifier

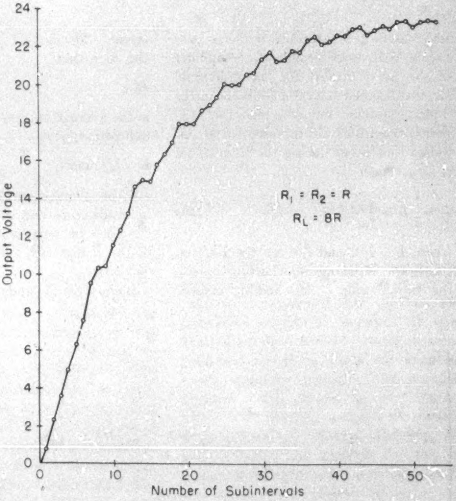


Fig. 13. Unit step response of 3-phase amplifier

In the analysis of the single-phase series-connected saturable inductor, a transformation of variables, physically interpreted as the action of a full-wave rectifier connected to the output, reduced the number of equations describing the operation of the amplifier from four to two. It was pointed out that this reduction was possible, because the action of the full-wave rectifier rendered the circuit symmetric so that no distinction could be made between the two intervals of operation. The same procedure can be applied to the 3-phase amplifier, and the addition of a full-wave rectifier at the output results in the transformation

$$\bar{I} = D\bar{I}' \qquad (153)$$

and

$$\bar{E} = D\bar{E}' \qquad (154)$$

where D is a 12×12 diagonal matrix in which the 8th, 10th, and 12th diagonal elements are -1, and the remaining elements are $+1$. (The first six elements of the $\bar{I}$ and $\bar{E}$ matrices are input quantities and are not changed; the second six are output quantities for the periods c_1, c_2, c_3, g_1, g_2, g_3 and are reversed in alternate periods.) This transformation changes the signs of the matrix coefficients as is shown by the signs in parentheses in Fig. 12. Addition of the six first rows and the six last rows after the transformation is made and substitution of equations 141 through 144 for the constants a, b, c, and d yield the admittance matrix equation:

$$\begin{Vmatrix} \bar{I}_1 \\ \bar{I}_2 \end{Vmatrix} = Y \begin{Vmatrix} \bar{E}_1 \\ \bar{E}_2 \end{Vmatrix} \qquad (155)$$

where Y is the matrix

$$Y = \begin{Vmatrix} \dfrac{1+z^{-1/3}+z^{-2/3}+z^{-1}+z^{-4/3}+z^{-5/3}}{(R_1+R_2)+(R_1-R_2)/z} & -\dfrac{1-z^{-1/3}+z^{-2/3}-z^{-1}+z^{-4/3}-z^{-5/3}}{(R_i+R_2)+(R_1-R_2)/z} \\ -\dfrac{1+z^{-1/3}+z^{-2/3}+z^{-1}+z^{-4/3}+z^{-5/3}}{(R_1+R_2)+(R_1-R_2)/z} & \dfrac{1-z^{-1/3}+z^{-2/3}-z^{-1}+z^{-4/3}-z^{-5/3}}{(R_1+R_2)+(R_1-R_2)/z} \end{Vmatrix} \qquad (156)$$

and

$$\bar{I}_1 = (\bar{I}')_1^{c_1} + (\bar{I}')_1^{c_2} + (\bar{I}')_1^{c_3} + (\bar{I}')_1^{g_1} + (\bar{I}')_1^{g_2} + (\bar{I}')_1^{g_3} \qquad (157)$$

and similarly for $\bar{I}_2$, $\bar{E}_1$, and $\bar{E}_2$. If saturation of the inductors occurs within the specified subintervals, equation 155 completely describes the behavior of the 3-phase amplifier.

To determine the transfer characteristics when resistively loaded, the current $\bar{I}_2$ is eliminated from equation 155 by using the relation $\bar{I}_2 = -(1/R_L)\bar{E}_2$. Then, by taking the inverse Laplace transform of equation 155, the following difference equation is obtained.

$$E_2(n)=\frac{R_L}{R_1+R_2+R_L}[E_1(n)+E_1(n-1/3)+E_1(n-2/3)+E_1(n-1)+E_1(n-4/3)+E_1(n-5/3)+E_2(n-1/3)-E_2(n-2/3)+\left(1-\frac{R_1-R_2}{R_L}\right)E_2(n-1)-E_2(n-4/3)+E_2(n-5/3)] \quad (158)$$

To illustrate the dynamic behavior of the amplifier, the response to a unit step input for the case when $R_1=R_2=1/8\ R_L$ is shown in Fig. 13.

Conclusion

A general and organized method of magnetic-amplifier analysis has been developed. Although this method of analysis may not be the shortest one for a particular circuit, it provides a general means for analysis and gives a clearer insight to the over-all properties of magnetic amplifiers. Magnetic amplifiers with diodes can also be analyzed using the same method, as is demonstrated in reference 2. In particular, this method of analysis is useful in determining general properties of magnetic amplifiers.

Appendix. Matrix Relations for the Basic Magnetic-Amplifier Element

The matrices of equations 43 through 46 are as follows:

$$Z=\begin{Vmatrix} R_1+R_m & 0 & R_m & 0 \\ -R_m/s & R_1 & -R_m/s & 0 \\ R_m & 0 & R_2+R_m & 0 \\ -R_m/s & 0 & -R_m/s & R_2 \end{Vmatrix} \quad (159)$$

$$H=\begin{Vmatrix} R_1+R_3 & 0 & R_m/(R_2+R_m) & 0 \\ -R_3/s & R_1 & -[R_m/(R_2+R_m)]/s & 0 \\ -R_3/R_2 & 0 & 1/(R_2+R_m) & 0 \\ [R_3/R_2]/s & 0 & [R_m/(R_2+R_m)]/R_2s & 1/R_2 \end{Vmatrix} \quad (160)$$

where

$$R_3=R_mR_2/(R_m+R_2) \quad (161)$$

$$G=\begin{Vmatrix} 1/(R_1+R_m) & 0 & R_4/R_1 & 0 \\ R_m/R_1(R_1+R_m)s & 1/R_1 & -R_4/R_1s & 0 \\ R_4/(R_1+R_4) & 0 & R_2+R_4 & 0 \\ -R_4/(R_1+R_4)s & 0 & -R_4/s & R_2 \end{Vmatrix} \quad (162)$$

where

$$R_4=R_mR_1/(R_m+R_1) \quad (163)$$

The elements of the A, B, C, and D matrix are given by the following expressions:

$$A=\begin{Vmatrix} (R_m+R_1)R_m & 0 \\ a/s & 0 \end{Vmatrix} \quad (164)$$

$$B=\begin{Vmatrix} (R_1R_m+R_2R_m+R_1R_2)/R_m & 0 \\ aR_2/s & 0 \end{Vmatrix} \quad (165)$$

$$C=\begin{Vmatrix} 1/R_m & 0 \\ (1+a)/R_1s & 0 \end{Vmatrix} \quad (166)$$

$$D=\begin{Vmatrix} (R_m+R_1)/R_m & 0 \\ R_2(1+a)/R_1s & 0 \end{Vmatrix} \quad (167)$$

where a is an arbitrary constant. The above expressions for A, B, C, and D were determined by solving the original equations. It was found that any value of the constant a will satisfy these equations. The arbitrary nature of a is a result of the property that port voltages and currents during the gating period do not affect port voltages or currents during the control period.

References

1. Impossible Behavior of Nonlinear Networks, R. J. Duffin. *Journal of Applied Physics*, New York, N. Y., vol. 26, 1955, pp. 603–05.

2. Analysis of Magnetic Amplifiers, P. R. Johannessen. *Report 7848-R-2*, Servomechanisms Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., Oct. 1958.

3. Communication Networks, Vol. II (book), E. A. Guillemin. John Wiley & Sons, Inc., New York, N. Y., 1935, pp. 145–50.

4. Analysis of Magnetic Amplifiers by the Use of Difference Equations, P. R. Johannessen. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 73, 1954 (Jan. 1955 section), pp. 700–11.

Analysis of Magnetic Amplifiers With Diodes

PAUL R. JOHANNESSEN
ASSOCIATE MEMBER AIEE

IT HAS BEEN demonstrated[1] that the behavior of magnetic amplifiers without diodes can be analytically described by a set of linear difference equations relating the average values of voltages and currents at the ports. For most circuits this representation gives values which are accurate to within 1 or 2% of actual even for a large variety of waveforms of port and supply voltages. The reason for this rather remarkable accuracy of representation can be attributed to three factors:

1. The control circuit resistance R_1, the output circuit resistance R_2, and the load resistance R_L are small compared to the magnetizing resistance R_m.
2. Almost all practical amplifiers of this type employ an interconnection of basic elements such that quiescent currents tend to cancel at the ports.
3. The state of the saturable inductor, saturated or unsaturated, depends on the integral of the inductor voltage rather than the inductor voltage itself.

The first factor insures that the saturable-inductor model is a good approximation to the actual saturable inductor. For all diodeless amplifiers the magnetizing resistance R_M can be considered to be infinite; thus, the ideal saturable-inductor model is sufficient to represent an actual inductor. The second factor insures that changes in the half-period average values of the supply voltage, caused by a shift in the operating period relative to the supply-voltage period, do not cause any steady-state errors in the linear analytic representation. However, the dynamic properties, as predicted by the linear representation, are affected by changes in the half-period average values of the supply voltage. The third

---

Paper 59-172, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted August 15, 1958; made available for printing January 12, 1959.

Paul R. Johannessen is with the Applied Research Laboratory of the Sylvania Electronic System, Waltham, Mass.

The material presented in this paper represents a portion of the thesis submitted by the author to the Massachusetts Institute of Technology in partial fulfillment of the requirements for the Doctor of Science degree. It was supported by the U. S. Air Force under Contract No. *AF(616)-5489*, MIT Project No. *DSR 7848*.

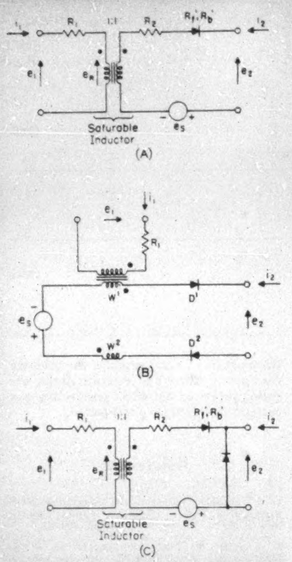


Fig. 1. Basic magnetic amplifiers

factor permits a large variety of waveforms of supply and port voltages to be used without effecting the accuracy of representation.

Magnetic amplifiers with diodes do not possess the aforementioned properties. The back impedance of the diode and the control circuit impedance are usually of the same order of magnitude as the magnetizing impedance. The accuracy of the saturable-inductor model is thus of paramount importance in designing a good analytical model. Cancellation of quiescent currents at the ports caused by interconnection of basic elements is also a property of magnetic amplifiers with diodes; this, however, does not necessarily insure linear steady-state characteristics. Probably the most important factor which distinguishes the behavior of these two types of magnetic amplifiers from each other is the manner in which their operation depends on voltage waveforms. The state of a diode depends upon the instantaneous voltage or current magnitudes while the state of a saturable inductor depends upon the integral of the inductor voltage, which, in turn, may depend upon the state of one diode or the state of several diodes, as well as the supply and port voltages. This interrelation greatly complicates the analysis and necessitates a complete reevaluation of the behavior of the basic elements.

Magnetic amplifiers with diodes invariably consist of symmetrical interconnections of basic elements. Basic elements typical for these types of magnetic amplifiers are shown in Fig. 1. In this paper only element a, of Fig. 1, is analyzed in detail since it is the most common one. Element (B), of Fig. 1, differs from element (A) only in regard to the physical arrangement of components. Thus, the analysis to be performed of element (A) applies equally well to element (B).

Since the objective of this paper is to demonstrate a new method of magnetic-amplifier analysis, the analysis is restricted to magnetic amplifiers terminated in a resistive load, and to voltage waveforms typical of such amplifiers. In the course of the analysis the restrictions upon the voltage waveforms that lead to a linear analysis are established. The voltage waveforms that satisfy these restrictions are termed ideal.

Analysis of Basic Magnetic Amplifier

The basic magnetic amplifier is shown in Fig. 1(A). The diode characteristic is approximated by a piecewise linear curve as shown in Fig. 2, where R_1' is the impedance of the diode when it is conducting and R_b' is the impedance of the diode when it is nonconducting. Two different models of the saturable inductor will be used: one to determine quiescent operating conditions and another to represent the saturable inductor in an incremental linear equivalent circuit. The inductor model which is used to determine the quiescent operating conditions is shown in Fig. 3. The finite loop-width property of the saturable inductor is taken into account by the diode bridge circuit and the constant current source. The diodes in the bridge circuit are assumed to be ideal; that is, having infinite conductance in the forward direction and zero conductance in the reverse direction. In Fig. 4, the complete equivalent circuit of the basic magnetic amplifier to be used in determining the quiescent operating conditions is shown. Use is made of the inequality $R_b' >> R_f'$ in Fig. 4 so that

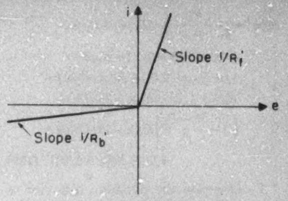


Fig. 2. Diode characteristic

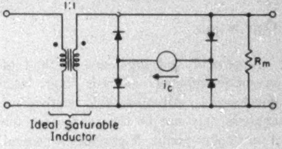


Fig. 3. Saturable inductor model

$$R_b \cong R_b' + R_2$$

$$R_f \cong R_f' + R_2 \qquad (1)$$

States of the Basic Magnetic Amplifier

The possible states of the equivalent circuit of Fig. 4 are shown in Table I. Note that each of the three nonlinear elements in Fig. 4 has two possible states. Combination of these states, therefore, should give rise to eight possible states of the basic magnetic amplifier. However, when the inductor is saturated, the state of the diode bridge is immaterial so that only six different combinations of states exist.

Typical Voltage Waveform

The voltage waveforms typical for magnetic amplifiers terminated in a resistive load are shown in Fig. 5. The voltage e_s passes through zero at time zero with positive slope, is symmetric about the $e=0$ axis, and is periodic of period T. The following approximations can then be made.

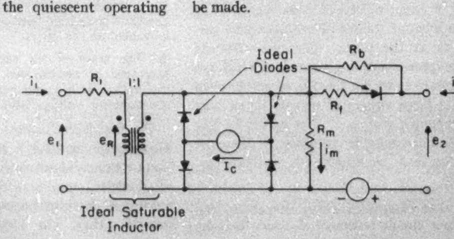


Fig. 4. Equivalent circuit of basic magnetic amplifier

Table I

State	Inductor: Unsaturated	Inductor: Saturated	Diode Bridge: Conductor	Diode Bridge: Nonconductor	Bridge: Conductor	Bridge: Nonconductor	Inductor Voltage
1		1			1		0
2		1				1	0
3	1			1	1		
4	1			1		1	
5	1		1			1	0
6	1		1		1		0

$$e_s=\frac{E_s}{T_s}t,\ \text{for } -\Delta t<t<\Delta t \tag{2}$$

$$e_s=-\frac{E_s}{T_s}\left(t-\frac{T}{2}\right),\ \text{for } \frac{T}{2}-\Delta t<t<\frac{T}{2}+\Delta t \tag{3}$$

$$e_s=\frac{E_s}{T_s}(t-T),\ \text{for } T-\Delta t<t<T+\Delta t \tag{4}$$

where

E_s = the half-period average value of the supply voltage
T_s = constant governed by the slope of e_s at a zero crossing

Note that the ratio E_s/T_s is the slope of the e_s wave at a zero crossing. For the same time intervals, the input and output waveforms can be approximated as follows:

$$e_1=E_1'+\frac{E_1}{T_1}t,\ \text{for } -\Delta t<t<\Delta t \tag{5}$$

$$e_1=E_1''-\frac{E_1}{T_1'}\left(t-\frac{T}{2}\right),\ \text{for } \frac{T}{2}-\Delta t<t<\frac{T}{2}+\Delta t \tag{6}$$

$$e_1=E_1'''+\frac{E_1}{T_1''}(t-T),\ \text{for } T-\Delta t<t<\frac{T}{2}+\Delta t \tag{7}$$

and

$$e_2=-\frac{E_2}{T_2}t,\ \text{for } -\Delta t<t<0$$

$$e_2=0,\ \text{for } 0<t<\Delta t \tag{8}$$

$$e_2=\frac{E_2}{T_2}\left(t-\frac{T}{2}\right),\ \text{for } \frac{T}{2}-\Delta t<t<\frac{T}{2}$$

$$e_2=0,\ \text{for } \frac{T}{2}<t<\frac{T}{2}+\Delta t \tag{9}$$

$$e_2=-\frac{E_2}{T_2}(t-T),\ \text{for } T-\Delta t<t<0$$

$$e_2=0,\ \text{for } T<t<T+\Delta t \tag{10}$$

where

E_1', E_1'', and E_1''' = the voltage magnitudes of e_1 at time 0, $T/2$, and T
E_1/T_1, E_1/T_1', and E_1/T_1'' = the magnitudes of the slope of e_1 at time 0, $T/2$, and T
E_2/T_2 = the magnitude of the slope of e_2 just before a zero crossing of the e_s wave

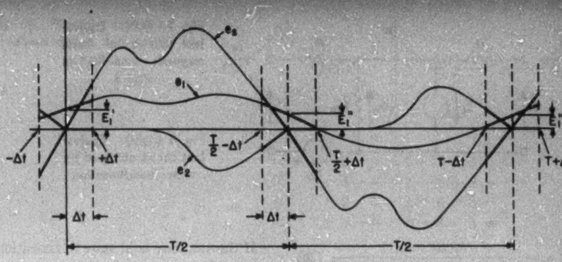


Fig. 5. Voltage waveforms

For reasons of analytical simplicity, all transitions from one state to another are assumed to occur within the time intervals in which the waveform approximations are made. Suppose that at some time t_{-1} in the interval $-\Delta t$ to 0 the inductor is saturated, the diode is conducting, and the operating point P on the inductor characteristic is at point P_{-1} in Fig. 6; saturation thus continues and the inductor voltage, e_R, is zero until the inductor current is equal to I_c. For $t=t_{-1}$, the basic magnetic amplifier is in state 1. In Fig. 6, the state of the amplifier is indicated by the superscripts on the P's. For $t>t_{-1}$, a transition occurs to either state 2 or 3. Transition to state 2 occurs if at some time $t_0>t_{-1}$

$$e_1-R_1I_c<0 \tag{11}$$

$$e_s+e_2=0 \tag{12}$$

For the waveforms chosen it is evident that $t_0=0$. Substitution of the waveform approximation for e_1 in equation 11 yields

$$E_1'<R_1I_c \tag{13}$$

Note that in state 2 the inductor voltage is zero.

If, instead, transition from state 1 to state 3 occurs at some time $t_0'>t_{-1}$, then the following conditions must be satisfied at time $t=t_0'$.

$$e_s+e_2=-\frac{R_f}{R_1}(e_1-R_1I_c) \tag{14}$$

$$e_s+e_2<0 \tag{15}$$

Substitution of equation 15 in equation 14 yields

$$e_1-R_1I_c>0,\ \text{for } t=t_0' \tag{16}$$

or in terms of the approximate voltage waveforms

$$E_1'+\frac{E_1}{T_1}t_0'>R_1I_c \tag{17}$$

As seen from Fig. 5, e_2 is assumed to be less than e_s. This assumption, together with equation 15, leads to a negative value for the time t_0',

$$t_0'<0 \tag{18}$$

From equations 17 and 18

$$E_1'>R_1I_c \tag{19}$$

An expression for t_0' in terms of the approximate waveforms is

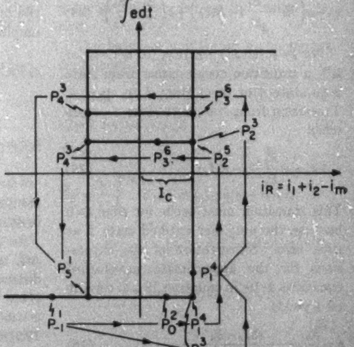


Fig. 6. Inductor characteristic

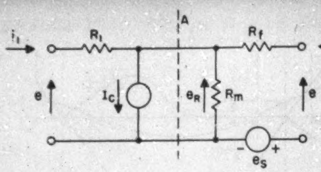
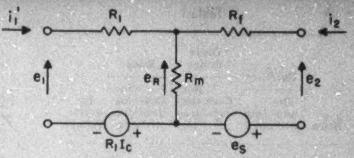


Fig. 7 (left). Equivalent circuit of basic magnetic amplifier in state 3

Fig. 8 (right). Equivalent circuit obtained by source transformation



$$t_0'=-\frac{R_f/R_1(E_1'-R_1I_c)}{E_s/T_s-E_2/T_2+(R_f/R_1)(E_1/T_1)} \tag{20}$$

When the amplifier is in state 3, it can be represented by the equivalent circuit of Fig. 7. By use of this equivalent circuit, the inductor voltage can be found in the following manner. If the section of the equivalent circuit on the left side of cut A is transformed into its Thevenin equivalent, the circuit in Fig. 8 results. By inspection of Fig. 8, equation 21 for the inductor voltage is obtained.

$$e_R=a(e_1-R_1I_c)+b(e_s+e_2) \tag{21}$$

where

a = the voltage transfer ratio between the e_1 (or $-R_1I_c$) and e_R (or e_1) terminals (measured with $e_2=e_s=R_1I_c=0$) when the diode is conducting

b = the voltage transfer ratio between the e_s (or e_2) and e_R terminals (measured with e_2 (or e_s) $=e_1=R_1I_c$ (or $e_1-R_1I_c$) $=0$) when the diode is conducting

In terms of the circuit parameters, the expressions for a and b are

$$a=\frac{R_mR_f}{R_1R_m+R_1R_f+R_mR_f} \tag{22}$$

$$b=\frac{R_mR_1}{R_1R_m+R_1R_f+R_mR_f}=a\frac{R_1}{R_f} \tag{23}$$

In terms of the approximate waveforms the expression for the inductor voltage then becomes

$$e_R=a\left(E_1'+\frac{E_1}{T_1}t-R_1I_c\right)+b\left(\frac{E_s}{T_s}-\frac{E_2}{T_2}\right)t \tag{24}$$

For $t=t_1$ (or t_1') where t_1 (or t_1')$>t_0$ (or t_0'), a transition occurs either from state 2 to state 4 or from state 3 to state 4. Transition from state 2 to state 4 occurs when

$$e_s+e_2=-\frac{R_b}{R_1}(e_1-R_1I_c) \tag{25}$$

This transition must occur for time $t>0$ because the amplifier entered state 2 at time zero. Substitution of the expressions for the approximate waveforms, equations 2–10, in equation 25 and solving for t_1 yields

$$t_1=\frac{R_b}{R_1}\frac{R_1I_c-E_1'}{E_s/T_s+(R_b/R_1)(E_1/T_1)} \tag{26}$$

If the amplifier is in state 3, transition to state 4 occurs at time t_1' when

$$e_s+e_2=\frac{R_m}{R_1+R_m}(e_1-R_1I_c) \tag{27}$$

In state 3 the inductor voltage is positive. Thus, the instant of time t_1' must occur after $t=0$. Substituting the approximate expressions for the voltage waveforms in equation 27 and solving for t_1' gives

$$t_1'=\frac{R_m}{R_1+R_m}\frac{E_1'-R_1I_c}{E_s/T_s-\dfrac{R_m}{R_1+R_m}E_1/T_1} \tag{28}$$

The equivalent circuit of the amplifier in state 4 is the same as that of Fig. 8, state 3, except that the diode resistance is R_b instead of R_f. The inductor voltage in state 4 thus becomes

$$e_r=c(e_1-R_1I_c)+d(e_s+e_2) \tag{29}$$

where the constants c and d are voltage transfer ratios similar to a and b in equation 21 except that the diode in nonconducting instead of conducting. In terms of the circuit parameters expressions for c and d are

$$c=\frac{R_mR_b}{R_1R_m+R_1R_b+R_mR_b} \tag{30}$$

$$d=\frac{R_mR_1}{R_1R_m+R_1R_b+R_mR_b}=c\frac{R_1}{R_b} \tag{31}$$

The transitions discussed in the foregoing are assumed to take place within the interval $-\Delta t$ to Δt. At time Δt the amplifier is in state 4. If the conditions

$$e_s+e_2>-\frac{R_b}{R_1}(e_1-R_1I_c),\ \text{for}\ \Delta t<t<\frac{T}{2}-\Delta t \tag{32}$$

$$e_s+e_2>\frac{R_m}{R_1+R_m}(e_1-R_1I_c),\ \text{for}\ \Delta t<t<\frac{T}{2}-\Delta t \tag{33}$$

are satisfied, then the amplifier remains in state 4.

In the interval $(T/2-\Delta t)$ to $(T/2+\Delta t)$, transitions similar to those already discussed for the interval $-\Delta t$ to Δt take place. If the inequality of equation 32 is violated in the interval $(T/2-\Delta t)$ to $T/2$, then transition to state 5 occurs at time

$$t_2=\frac{T}{2}-\frac{R_b}{R_1}\frac{R_1I_c-E_1''}{E_s/T_s-E_2/T_2+(R_b/R_1)(E_1/T_1')} \tag{34}$$

Note that state 5 and state 1 are equivalent in as far as current and voltage relations are concerned. If the inequality of equation 33 is violated, then transition to state 3 occurs at time

$$t_2'=\frac{T}{2}-\frac{R_m}{R_1+R_m}\times\frac{E_1''-R_1I_c}{E_s/T_s-E_2/T_2-\dfrac{R_m}{R_1+R_m}E_1/T_1'} \tag{35}$$

Note that t_2' must occur before $T/2$. In state 3 the inductor voltage is given by equation 21.

Next, transition from state 5 or 3 to state 6 occurs. If the amplifier is in state 5, then transition to state 6 occurs at time

$$t_3=T/2 \tag{36}$$

If the waveform of e_1 had been chosen so that E_1'' is negative, then the transition from state 5 to state 6 may occur after time $T/2$. This event takes place if $-E_1''>R_1I_c$. Thus, if $-R_1I_c<E_1''<R_1I_c$, then transition to state 6 occurs at time $t_3=T/2$.

If the amplifier is in state 3, then transition to state 6 occurs when

$$e_s+e_2=-\frac{R_f}{R_1}(e_1-R_1I_c) \tag{37}$$

In state 3 the diode is conducting; the inductor voltage must then be greater than e_s+e_2. Thus the time of transition from state 3 to state 6 occurs after time $T/2$, and is given by the expression

$$t_3'=T/2+\frac{R_f}{R_1}\frac{E_1''-R_1I_c}{E_s/T_s+(R_f/R_1)(E_1/T_1')} \tag{38}$$

The next transition is from state 6 to state 3 and occurs when

$$e_1+e_2=-\frac{R_f}{R_1}(e_1+R_1I_c) \tag{39}$$

and at time

$$t_4=T/2+\frac{R_f}{R_1}\frac{E_1''+R_1I_c}{E_s/T_s+(R_f/R_1)(E_1/T_1')} \tag{40}$$

The inductor voltage for state 3 is supposedly given by equation 21. This equation was derived for the operating point P_0^3 in Fig. 6. For the present case the operating point is on the other side of the inductor characteristic. Thus, for the present case, the current source I_c in the equivalent circuit of Fig. 7 is of opposite polarity so that the term $e_1 - R_1 I_c$ in the equation for the inductor voltage becomes $e_1 + R_1 I_c$. For the amplifier to remain in state 3 until the inductor saturates at t_6, $e_s + e_2$ must satisfy both of the following inequalities

$$e_s + e_2 < -\frac{R_f}{R_1}(e_1 + R_1 I_c) \tag{41}$$

$$e_s + e_2 < \frac{R_m}{R_1 + R_m}(e_1 + R_1 I_c) \tag{42}$$

Equation 41 insures that the net inductor current is greater than I_c, while equation 42 insures that the diode conducts. For the waveforms of Fig. 5, equation 41 is satisfied if equation 42 is satisfied. When the inductor saturates at time t_6, transition to state 1 occurs and saturation continues until a new period of operation starts. It is evident that the sequence of transitions taking place in the next period of operation may not be the same as in the period just considered. However, if all voltage waveforms are periodic, then the sequence of transitions are the same for all periods of operation.

A diagram illustrating the possible sequences of transitions that may occur during one period of operation is shown in Fig. 9. These sequences of transitions are the only ones that can occur for voltage waveforms of the type shown in Fig. 5 that satisfy equations 32, 33, 41, and 42. Clearly, for voltage waveforms different from those shown in Fig. 5 that may or may not satisfy equations 32, 33, 41, and 42, there exists a large number of possible sequences of transitions. It is beyond the scope of this work to consider all of these possibilities.

The waveforms of Fig. 5 were chosen in order to demonstrate a method of analysis and to establish restrictions upon the voltage waveforms that render the equations for the basic amplifier linear. To establish these waveform restrictions, the amplifier equations for the four sequences of transitions in Fig. 9 must be studied. In Appendix I the amplifier equations for sequence of transitions c are derived, and it is shown that these equations are linear if either of the two equations, 163 and 164, together with equations 32, 33, and 42, are satisfied for all periods of operations.

The conditions of equation 164 may appear to be trivial, since they imply that all the voltage waves are constant near the zero crossings of the e_s wave. The author, however, has not been able to derive the linearization conditions of equations 163 and 164 without going through the detailed study of possible sequences of transitions as given in this section and the derivation of the amplifier equations as presented in Appendix I. It is further shown in Appendix I that if these waveform restrictions are satisfied, then the amplifier equations are linear for all four sequences of transitions.

Equations 163 and 164 are not the only linearization conditions that exist. A general study of all possible relations between the various voltage waveforms that render the amplifier equations linear appears to be of little practical value. The conditions of equations 163 and 164, however, are often met in practical amplifiers. The linearization condition of equation 163 is of special interest since considerable freedom is allowed for the various voltage waveforms. This condition, together with the conditions of equations 32, 33, and 42, may be formulated in terms of four alternate conditions as follows:

1. The operation of the amplifier must be periodic and of the same period as the supply voltage.
2. The diode must be nonconducting during the control periods and conducting during the gating periods.
3. The inductor voltage must be positive during the control periods.
4. The input voltage must satisfy the relation.

$$e_1 = R_1 I_c, \text{ for } t = T/2 \tag{43}$$

These conditions need not be satisfied in order to perform a linear, small-signal analysis around the quiescent operating point. This type of analysis is most easily performed by first determining the quiescent operating conditions, as outlined in Appendix I for the combination of transitions c in Fig. 9, and then linearizing around the quiescent operating point.

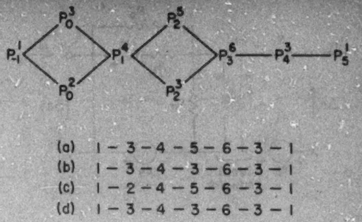

P_{-1}^1
P_0^3
P_0^2
P_1^4
P_2^5
P_2^3
P_3^6
P_4^3
P_5^1
(a) 1 − 3 − 4 − 5 − 6 − 3 − 1
(b) 1 − 3 − 4 − 3 − 6 − 3 − 1
(c) 1 − 2 − 4 − 5 − 6 − 3 − 1
(d) 1 − 3 − 4 − 3 − 6 − 3 − 1

Fig. 9. Possible sequences of transitions during one period of operation

CIRCUIT EQUATIONS FOR BASIC MAGNETIC AMPLIFIERS THAT SATISFY THE FIRST THREE LINEARIZATION CONDITIONS

In basic magnetic amplifiers that satisfy conditions 1, 2, and 3, there exist two possible sequences of transitions:

(1) 1−4−5−6−3−1

(2) 1−4−6−3−1

Only sequence 1 need be considered because sequence 2 is actually a special case of 1 for which the time spent in state 5 zero.

During the control period, the following equations hold,

$$E_1^c = R_1 I_1^c + E_R^c \tag{44}$$

$$E_s^c + E_2^c = R_b I_2^c + E_R^c \tag{45}$$

and during the gating period

$$E_1^g = R_1 I_1^g + E_R^g \tag{46}$$

$$E_s^g + E_2^g = R_f I_2^g + E_R^g \tag{47}$$

For the foregoing, the notation is the same as used in reference 1. To obtain four equations involving the eight port voltages and currents only, two more equations are needed. One of these equations is the well-known relation

$$E_R^c + E_R^g = 0 \tag{48}$$

the other is obtained by calculating the inductor voltage during the control period.

The inductor voltage is given by the expressions

$$e_R = c(e_1 - R_1 I_c) + d(e_s + e_2), \text{ for } 0 < t < t_2 \tag{49}$$

$$e_R = 0, \text{ for } t_2 < t < T/2 \tag{50}$$

Thus,

$$\begin{aligned} E_R^c &= \frac{2}{T}\int_0^{t_2} [c(e_1 - R_1 I_c) + d(e_s + e_2)]\,dt \\ &= \frac{2}{T}\int_0^{T/2} [c(e_1 - R_1 I_c) + d(e_s + e_2)]\,dt - \\ &\quad \frac{2}{T}\int_{t_2}^{T/2} [c(e_1 - R_1 I_c) + d(e_s + e_2)]\,dt \end{aligned} \tag{51}$$

which, when integrated, gives

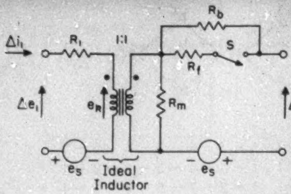


Fig. 10. Linear incremental model of basic magnetic amplifier

$$E_R{}^c = c(E_1{}^c - R_1I_c) + c\frac{R_1}{R_b}(E_s{}^c + E_2{}^c) + \frac{1}{T}\frac{R_b}{R_1}c\frac{1}{A_s - A_2 + (R_b/R_1)A_1{}'}(E_1{}'' - R_1I_c)^2 \qquad (52)$$

where

$A_s = E_s/T_s$ = the magnitude of the slope of the e_s wave at time $T/2$

$A_2 = E_2/T_2$ = the magnitude of the slope of the e_s wave immediately before time $T/2$

$A_1{}' = E_1/T_1{}'$ = the magnitude of the slope of the e_1 wave at time $T/2$

If the slopes A_s, A_2, and $A_1{}'$, and the voltage $E_1{}''$ are independent of the half-period average values of the voltages and currents at ports 1 and 2, and if the supply voltage is constant, then equation 52 can be rewritten

$$E_R{}^c = cE_1{}^c + d(E_s{}^c + E_2{}^c) - B \qquad (53)$$

where B is a constant given by the expression

$$B = cR_1I_c - \frac{1}{T}\frac{R_b}{R_1}c\frac{(E_1{}'' - R_1I_c)^2}{A_s - A_2 + (R_b/R_1)A_1{}'} \qquad (54)$$

If $E_1{}'' = R_1I_c$, then $E_R{}^c$ is independent of the slopes A_s, A_2, and $A_1{}'$, and the expression for B becomes simply

$$B = cR_1I_c \qquad (55)$$

This is the special case for which the sequence of transition 2 applies.

Substitution of equations 48 and 53 in equations 44, 45, 46, and 47, with an appropriate rearrangement of terms, yields the admittance-matrix equation of the amplifier.

$$\left\|\begin{matrix} I_1{}^c \\ I_1{}^g \\ I_2{}^c \\ I_2{}^g \end{matrix}\right\| = \left\|\begin{matrix} \frac{1-c}{R_1} & 0 & -\frac{d}{R_1} & 0 & -\frac{d}{R_1} & 0 \\ \frac{c}{R_1} & \frac{1}{R_1} & \frac{d}{R_1} & 0 & \frac{d}{R_1} & 0 \\ -\frac{c}{R_b} & 0 & \frac{1-d}{R_b} & 0 & \frac{1-d}{R_b} & 0 \\ \frac{c}{R_f} & 0 & \frac{d}{R_f} & \frac{1}{R_f} & \frac{d}{R_f} & \frac{1}{R_f} \end{matrix}\right\| \left\|\begin{matrix} E_1{}^c \\ E_1{}^g \\ E_2{}^c \\ E_2{}^g \\ E_s{}^c \\ E_s{}^g \end{matrix}\right\| + \beta \qquad (56)$$

where B is the column matrix

$$\beta = \left\|\begin{matrix} \frac{1}{R_1}B \\ -\frac{1}{R_1}B \\ -\frac{1}{R_b}B \\ \frac{1}{R_f}B \end{matrix}\right\| \qquad (57)$$

More generally, any desired network matrix can be obtained from these same equations.

Incremental Linear Analysis

Suppose that the half-period average voltages and currents of equation 56 are the quiescent currents and voltages, then the admittance-matrix equation which relates the Δ quantities at ports 1 and 2 is

$$\left\|\begin{matrix} \Delta I_1{}^c \\ \Delta I_1{}^g \\ \Delta I_2{}^c \\ \Delta I_2{}^g \end{matrix}\right\| = \left\|\begin{matrix} \frac{1-c}{R_1} & 0 & -\frac{d}{R_1} & 0 \\ \frac{c}{R_1} & \frac{1}{R_1} & \frac{d}{R_1} & 0 \\ -\frac{c}{R_b} & 0 & \frac{1-d}{R_b} & 0 \\ \frac{c}{R_f} & 0 & \frac{d}{R_f} & \frac{1}{R_f} \end{matrix}\right\| \left\|\begin{matrix} \Delta E_1{}^c \\ \Delta E_1{}^g \\ \Delta E_2{}^c \\ \Delta E_2{}^g \end{matrix}\right\| \qquad (58)$$

In general, this equation is valid only if the Δ quantities result from incremental changes in the waveforms of the quiescent currents and voltages. Thus, equation 58 is a linear relation among incremental deviations from the quiescent condition. It is of importance to point out that the restriction of small changes in the average values of the port quantities is not always necessary. If the changes in the waveforms are such that the total voltage waves at ports 1 and 2 satisfy the conditions for ideal waveforms, then the amplifier is linear and is represented by equation 58.

A linear-incremental circuit model of the basic magnetic amplifier is shown in Fig. 10. The switch S is open during the positive half periods and closed during the negative half periods of e_s.

In the preceding analysis, it was found that, in general, the amplifier equations are not linear. The nonlinearity in the equations is introduced by the last term of equation 52 and is caused by changes in the slopes of the voltage waves near the zero crossings of the e_s wave and the quantity $E_1{}''$ with applied signals. For practical amplifiers terminated in a resistive load, the slope of the output voltage wave near the zero crossings of the e_s wave is governed by the e_s wave only. This is true in general for single-phase amplifiers terminated in a resistive load; for 3-phase amplifiers, however, this is true only if the amplifier operates in mode 1 and is resistively loaded. If the e_s wave is assumed to be invarient, then the following analysis can be made. Let

$$A_q = A_{sq} - A_{2q} + (R_b/R_1)A_1{}'_q \qquad (59)$$

where the subscript q indicates the quiescent condition. Changes in $E_R{}^c$ caused by changes in $E_1{}^c$ and $E_2{}^c$ can then be written

$$\begin{aligned} \Delta E_R{}^c &= E_R{}^c{}_q + \Delta E_R{}^c - E_R{}^c{}_q \\ &= c(E_1{}^c{}_q + \Delta E_1{}^c - R_1I_c) + \\ &\quad d(E_s{}^c{}_q + E_2{}^c{}_q + \Delta E_2{}^c) + \\ &\quad \frac{1}{T}\frac{R_b}{R_1}c\frac{(E_1{}''{}_q + \Delta E_1{}'' - R_1I_c)^2}{A_q + \Delta A} - \\ &\quad c(E_1{}^c{}_q - R_1I_c) - d(E_s{}^c{}_q + E_2{}^c{}_q) - \\ &\quad \frac{1}{T}\frac{R_b}{R_1}c\frac{(E_1{}''{}_q - R_1I_c)^2}{A_q} \end{aligned} \qquad (60)$$

For small Δ changes the following approximation can be made

$$\frac{1}{A_q + \Delta A} \cong \frac{1}{A_q}\frac{1}{1 + \frac{\Delta A}{A_q}} = \frac{1}{A_q}\left(1 - \frac{\Delta A}{A_q}\right) \qquad (61)$$

Substitution of equation 61 in equation 60 and cancellation of all Δ^2 terms yield

$$\Delta E_R{}^c = c\Delta E_1{}^c + d\Delta E_2{}^c + \frac{1}{T}\frac{R_b}{R_1}c\cdot\frac{2}{A_q}(E_1{}''{}_q - R_1I_c)\Delta E_1{}'' - \frac{1}{T}c\left(\frac{R_b}{R_1}\right)^2\left(\frac{1}{A_q}\right)^2(E_1{}''{}_q - R_1I_c)^2\Delta A \qquad (62)$$

The quantity $\Delta E_1{}''$ is the change in the magnitude of the e_1 wave at time $T/2$. Usually, $\Delta E_1{}''$ is linearly related to $\Delta E_1{}^c$, $E_1{}''$ is equal to $E_1{}^c$ for direct current, and is zero for a full-wave rectified sine wave. Let

$$\Delta E_1{}'' = K_1\Delta E_1{}^c \qquad (63)$$

where K_1 is a constant.

The quantity ΔA is the change in the slope of the e_1 wave at time $T/2$. Clearly, if the waveform of e_1 is fixed and only the magnitude is varied, then

$$\Delta A=\frac{1}{T_1'}\Delta E_1^c \qquad (64)$$

where T_1' is a constant. Substitution of equations 63 and 64 in equation 62 yields

$$\Delta E_R^c=c'\Delta E_1^c+d\Delta E_2^c \qquad (65)$$

where

$$c'=c\left\{1+\frac{1}{T}\frac{R_b}{R_1}\frac{1}{A_q}(E_1''_q-R_1I_c)\times\right.$$
$$\left.\left[2K_1-\frac{1}{T_1'}\frac{R_b}{R_1}\frac{1}{A_q}(E_1''_q-R_1I_c)\right]\right\} \qquad (66)$$

Equation 65, together with equations 44, 45, 46, 47, and 48, leads again to a linear analysis of the basic magnetic amplifier in terms of incremental deviation from the quiescent condition.

It is evident from equation 66 that, for the special cases in which either A_q is infinite, that is, if either of the slopes A_{sq} and $A_1'_q$, or both, are infinite, or E_1'' is equal to R_1I_c, the quantity c' reduces to c.

The preceding considerations hold not only for incremental deviations from the quiescent condition but also for incremental deviations from any predetermined operating condition.

BIAS CONSIDERATIONS

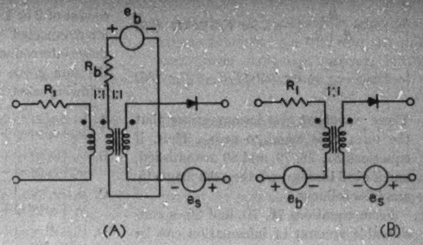
In most practical amplifiers a bias, or reset, circuit is provided to adjust the quiescent operating condition as desired. This is accomplished either by an additional bias winding on the inductor connected in series with a resistor and a voltage source or by a voltage source in series with the input. These two methods of obtaining the proper bias are shown in Fig. 11 and are referred to as parallel and series bias, respectively. The voltage gain of the amplifier is proportional to the component of the inductor voltage cE_1^c, where the voltage transfer ratio c is given by equation 30. In the case of parallel bias, the resistance R_b effectively shunts the inductor and thus decreases the value of the voltage transfer ratio c, which, in turn, decreases the voltage gain. It is thus desirable to make the resistance R_b as large as possible. In the limiting case when R_b approaches infinity, the resistance R_b and the voltage source e_b can be replaced by a constant current source i_b. This current source, by Thevinin's theorem, is equivalent to a voltage source of magnitude R_1i_b in series with the input voltage. Thus the two bias methods become identical and no loss in gain results. If

$$e_b=R_1I_c \qquad (67)$$

then the bias voltage cancels the equivalent voltage source R_1I_c during the control periods. Effectively, this bias voltage shifts the hysteresis loop in Fig. 6 to the left by an amount I_c so that the apparent loop width is zero during the control periods and equal to $2I_c$ during the gating periods.

Fig. 11. Bias methods

A—Parallel
B—Series



With a bias voltage, or current, in the amplifier, the ideal waveform restrictions are altered slightly. When the bias voltage is equal to R_1I_c, these restrictions become

$$E_1'=E_1''=E_1'''=0 \qquad (68)$$

$$e_s+e_2>\frac{R_m}{R_1+R_m}e_1,\ \text{for } 0<t<\frac{T}{2} \qquad (69)$$

$$e_s+e_2<\frac{R_m}{R_1+R_m}(e_1+2R_1I_c),\ \text{for } \frac{T}{2}<t<t_5 \qquad (70)$$

$$e_s+e_2>-\frac{R_b}{R_1}e_1,\ \text{for } 0<t<T/2 \qquad (71)$$

The expression for the term B in the equation for the inductor voltage, equation 53, becomes

$$B=-\frac{1}{T}\frac{R_b}{R_1}c\frac{(E_1'')^2}{A_s-A_2+(R_b/R_1)A_1'} \qquad (72)$$

From the relation

$$E_2^g=R_fI_2^g-E_R^c-E_s^g \qquad (73)$$

obtained by substituting equation 48 in equation 47, it is evident that the relation between the port quantities E_2^g and I_2^g does not depend upon the inductor characteristics during the gating periods. In fact, any finite resistance can be placed across the inductor during the gating periods without affecting the performance of the amplifier. In addition, if the bias voltage is equal to R_1I_c, the apparent loop width is equal to zero during the control periods. Thus, for ideal waveforms and the bias voltage of equation 67, the amplifier is linear and independent upon the constant loop-width property of the saturable inductor. For this reason, the bias voltage of equation 67 is termed the ideal.

A SET OF IDEAL VOLTAGE-WAVEFORM RESTRICTIONS

To study voltage waveforms in practical amplifiers, a set of sufficient, although not necessary, restrictions upon the voltage waveforms for the case when

$$e_s+e_2>0,\ \text{for } 0<t<\frac{T}{2}$$

$$e_s+e_2<0,\ \text{for } \frac{T}{2}<t<T \qquad (74)$$

are derived from the waveform restrictions of equations 68 through 71 as follows.

From equation 74,

$$e_s+e_2=|e_s+e_2|,\ \text{for } 0<t<\frac{T}{2} \qquad (75)$$

Thus, if

$$|e_s+e_2|>\frac{R_m}{R_1+R_m}e_1,\ \text{for } 0<t<\frac{T}{2} \qquad (76)$$

then equation 69 is satisfied.

Equation 70 can be alternatively written

$$-(e_s+e_2)>-\frac{R_m}{R_1+R_m}(e_1+2R_1I_c),$$
$$\text{for } \frac{T}{2}<t<t_5 \qquad (77)$$

From equation 74,

$$-(e_s+e_2)=|e_s+e_2|,\ \text{for } \frac{T}{2}<t<T \qquad (78)$$

Thus, if

$$|e_s+e_2|>\frac{R_m}{R_1+R_m}(-e_1),\ \text{for } \frac{T}{2}<t<t_5 \qquad (79)$$

then equation 70 is satisfied.

Through a similar derivation,

$$|e_s+e_2|>-\frac{R_b}{R_1}e_1,\ \text{for } 0<t<T/2 \qquad (80)$$

then equation 71 is satisfied. Thus, a set of sufficient, but not necessary, restrictons upon the voltage waveforms is

$$E_1'=E_1''=E_1'''=0 \qquad (81)$$

$$|e_s+e_2|>\frac{R_m}{R_1+R_m}e_1,\ \text{for } 0<t<T/2 \qquad (76)$$

$$|e_s+e_2| > \frac{R_m}{R_1+R_m}(-e_1), \text{ for } T/2<t<t_b \qquad (79)$$

$$|e_s+e_2| > -\frac{R_b}{R_1}e_1, \text{ for } 0<t<T/2 \qquad (80)$$

These waveform restrictions insure that the term B is equal to zero. Thus, if equations 81, 76, 79, and 80 are satisfied, and if ideal bias is used, then the magnetic amplifier is linear.

From equations 76, 79, and 80, a considerable amount of information can be obtained regarding the behavior of single-phase magnetic amplifiers terminated in a resistive load. Equation 76 essentially states that the maximum inductor voltage that can exist during the control periods without causing the diode to conduct is equal to the sum of the supply voltage and the output voltage. Because the inequality

$$R_1 >> R_f$$

usually is satisfied, the sum of the supply and output voltages establishes an upper limit on the inductor voltage during a control period. This limit, in turn, establishes an upper limit on the maximum flux reset that can occur during a control period. If equation 79 is satisfied, the diode is conducting and the component of the inductor voltage caused by the input voltage is small. Thus, the inductor is driven toward saturation by the supply and output voltages. However, if the input voltage is large and negative, such that equation 79 is not satisfied, then the diode is nonconducting and the input voltage is effective in driving the inductor toward saturation. In conclusion, it apparently is relatively easy to prevent the inductor from being reset during a control period and to drive the inductor toward saturation during a gating period. Difficulties may arise in attempting to reset the inductor during a control period and in preventing it from being driven toward saturation during a gating period. It may be said that the inductor "likes" to be saturated.

This property is basically responsible for several undesirable magnetic-amplifier limitations: the low-efficiency of all fast-response magnetic amplifiers; the impossibility of constructing a full-wave polarity-reversible d-c amplifier of greater than 50% efficiency; and the impossibility of constructing a fast-response, highly efficient magnetic amplifier with full-wave a-c output.

THE VOLTAGE e_s+e_2 IN PRACTICAL MAGNETIC AMPLIFIERS

A single-phase magnetic amplifier may consist of 2 or 4 basic magnetic amplifiers interconnected in various ways. For these interconnections the magnitude of the sum e_s+e_2 must satisfy one of the following cases.

A. $|e_s+e_2| \cong |e_s|+|e_2|$ control period
$|e_s+e_2| \cong |e_s|+|e_2|$ gating period

B. $|e_s+e_2| \cong |e_s|+|e_2|$ control period
$|e_s+e_2| \cong |e_s|-|e_2|$ gating period

C. $|e_s+e_2| \cong |e_s|-|e_2|$ control period
$|e_s+e_2| \cong |e_s|+|e_2|$ gating period

D. $|e_s+e_2| \cong |e_s|-|e_2|$ control period
$|e_s+e_2| \cong |e_s|-|e_2|$ gating period

E. $|e_s+e_2| \cong |e_s|$ control period
$|e_s+e_2| \cong |e_s|+|e_2|$ gating period

F. $|e_s+e_2| \cong |e_s|$ control period
$|e_s+e_2| \cong |e_s|-|e_2|$ gating period

G. $|e_s+e_2| \cong |e_s|$ control period
$|e_s+e_2| \cong |e_s| \pm |e_2|$ gating period

H. $|e_s+e_2| \cong |e_s| \pm |e_2|$ control period
$|e_s+e_2| \cong |e_s| \pm |e_2|$ gating period

A plus sign or a minus sign on the right-hand side indicates that addition or subtraction always occurs regardless of the polarity of the input voltage. A plus and a minus sign indicates that either addition or subtraction may occur, depending on the input voltage polarity. These cases represent the conditions that may exist for a single basic amplifier in a magnetic amplifier consisting of an interconnection of two or four basic amplifiers.

For cases A, B, C, and D, there is output during both half periods of operation. Amplifiers with this property are obtained by parallel connecting the output ports of two basic amplifiers that gate during alternate half periods. Inasmuch as output is obtained from the basic amplifier under consideration during its gating periods, the output voltage is always subtracted from the supply voltage during the gating periods. Cases A, C, and E, therefore, do not represent practical amplifiers. When the output adds to the supply voltage during the control periods and subtracts from it during the gating periods, the output voltage is always positive because the supply voltage is an alternating voltage. The output voltage is an alternating voltage when it subtracts from the supply voltage in both control and gating periods. Thus, the output for case B is a direct voltage; and for case D it is an alternating voltage. Case F represents the well-known incomplete bridge circuit shown in Fig. 12. This amplifier circuit may be constructed by connecting two basic amplifiers of the type shown in Fig. 1(C) in series-series. Case G represents the so-called fast response, or half-cycle response, magnetic amplifier. These amplifiers are constructed by paralleling the output ports of two basic amplifiers that gate during the same half periods. Case H represents the various full-wave phase-reversible, or polarity reversible, magnetic amplifiers. These amplifiers result from connecting the output ports of four basic amplifiers in parallel.

The ideal waveform restrictions are easily satisfied for the amplifiers belonging to cases B and F. These amplifiers can be completely reset during one-half period of the supply voltage because of the large diode blocking voltage. They are inherently efficient, and may be fast or slow response amplifiers; characteristically, they give a unipolar, direct-voltage output.

The amplifiers of case D are inherently slow-response high-gain amplifiers. Speed of response may be traded for lower efficiency and lower gain, for example, by connecting resistors in series with the diodes. These resistors decrease the output voltage and thus increase the diode blocking voltage. Amplifiers of this type are characterized by a full-wave alternating output voltage.

The fast response, bipolar (or phase reversible if the output is an alternating voltage) amplifiers of case *g* are characterized by a low efficiency, which is evident from the magnitude of the sum of the supply and output voltage. This voltage sum is smaller during the control periods than during the gating periods. It is, therefore, difficult to obtain sufficient reset during the control periods to prevent the inductor from saturating during the gating periods. Because the amplifier consists of two basic amplifiers that gate during the same half periods, a time interval exists in which both inductors are saturated; a large circulating current results giving rise to a low efficiency.

The amplifiers belonging to case H have properties common to all the aforementioned amplifiers. The various combinations of properties that these amplifiers may possess are left to the imagination of the reader.

Violation of the ideal voltage-waveform restrictions not only prevent a linear analysis from being made, but also result in undesirable properties such as low effi-

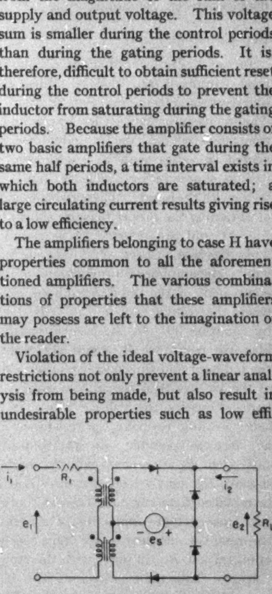


Fig. 12. Incomplete bridge circuit

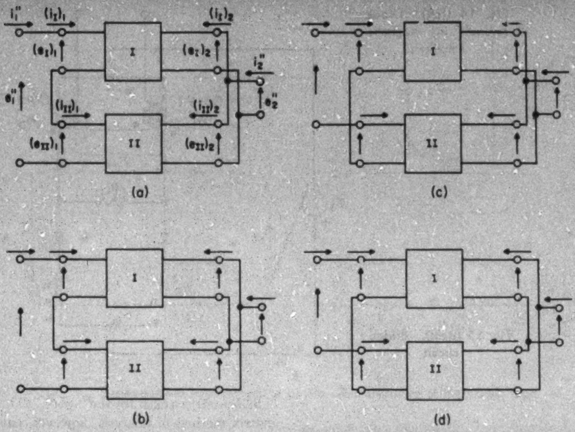


Fig. 13. The four possible subconnections for the series parallel interconnection of 2-port networks

ciency, low gain-bandwidth product, and nonlinear transfer characteristics.

Typical voltage waveforms that satisfy the ideal waveform restrictions for practical magnetic amplifiers are

1. Direct input voltage and rectangular supply voltage.
2. Alternating input voltage of the same waveform as the supply voltage.
3. Input voltage of the same waveform as half-wave or full-wave rectified supply voltage.
4. Input voltage of the same waveform as the waveform of the sum of the supply and output voltages.

NONIDEAL WAVEFORMS IN MAGNETIC AMPLIFIERS

The ideal voltage-waveform restrictions are formulated in terms of certain relations that must be satisfied among the various voltage waves. It is not logical to speak of an ideal supply, input, or output voltage, as such, because there are no restrictions imposed upon each of these voltages separately. If the supply voltage is a rectangular wave, then almost any reasonable input voltage wave will satisfy the ideal waveform restrictions. The combination of a sinusoidal supply voltage and a direct input voltage, however, does not satisfy the ideal waveform restrictions, because this combination of waveforms violates the restrictions set forth by equation 81. The amplifier nonlinearities resulting from these voltage waveforms are usually not very serious, as will be demonstrated in a later section.

The most frequently encountered input voltages which violate the ideal waveform restrictions and thus deteriorate the amplifier performance are high-frequency noise and, for amplifiers controlled by an alternating signal, quadrature signals. Both of these types of input voltages are of particular importance for the fast-response phase-reversible magnetic amplifiers. Such amplifiers are used extensively in feedback control systems in which subtractions of two large, almost equal signals are a common operation, the signal-to-noise ratio of the resultant signal being often as low as 0.01. Thus, properties such as high-frequency noise and quadrature rejection are highly desirable. Ideally, the magnetic amplifier is an average proportional amplifier; because the average value of high-frequency noise, or of random noise if the time interval in which the averaging process takes place is sufficiently long, and quadrature voltages are zero, under ideal operating conditions the magnetic amplifier should be insensitive to these input signals.

It has been shown previously that waveforms that violate the ideal-waveform restrictions tend to cause the inductor to saturate earlier in the gating period than do ideal voltage waveforms of the same half-period average values. This phenomenon causes the quiescent currents to increase in each of the basic amplifiers of which the magnetic amplifier is constructed. The quiescent currents may still cancel at the output, but the quiescent power dissipation increases, which, in turn, decreases the amplifier efficiency. This change in the quiescent operating condition with input noise is quite often a serious problem; not only does the quiescent power dissipation increase, but the amplifier may be driven to cutoff.

Interconnection of Basic Magnetic Amplifiers

In the previous section it was shown that, if the port and supply-voltage waveforms satisfy certain restrictions, the basic magnetic amplifier can be represented by a set of linear equations among average values of port voltages and currents. It was further shown that, even if these waveform restrictions are not satisfied a linear analysis can be performed in terms of incremental deviations from the quiescent condition.

This section is concerned with the analysis of magnetic amplifiers resulting from various interconnections of basic magnetic amplifiers whose port voltages and currents are linearly related.

INTERCONNECTION OF TWO BASIC AMPLIFIERS

As demonstrated in reference 1, there exist 4 different magnetic-amplifier circuits without diodes for each of the 5 basic interconnections. For magnetic amplifiers with diodes there exist 8 different circuits for each of the 5 basic interconnections. Consider, for example, subconnection *b* of the basic series-parallel interconnection of Fig. 13; the amplifier resulting is shown in Fig. 14(A). This amplifier belongs to case G of the preceding section since both inductors saturate during the same half period and the output is a half-wave polarity-reversible voltage. Suppose, now, that the diode of basic amplifier *II*, as seen in Fig. 13, is reversed. The resultant circuit is shown in Fig. 14(B), and it belongs to case B; the inductors saturate during alternate half periods and the output is a positive direct voltage. Thus, for each of the subconnections of Fig. 13, there exist 2 different circuits: one is obtained by connecting the ports of 2 ordinary basic amplifiers according to the particular subconnection and the other by reversing either one of the diodes in this circuit. It is beyond the scope of this work to consider all of the possible interconnections of 2 basic amplifiers; it must suffice to say that there exists a total of 40 different circuits inasmuch as there are 5 basic interconnections comprising 8 subconnections each. Not all of these interconnections result in practical amplifiers; for example, the circuits resulting from series connecting the output ports are impractical. However, by connecting a diode across the output port

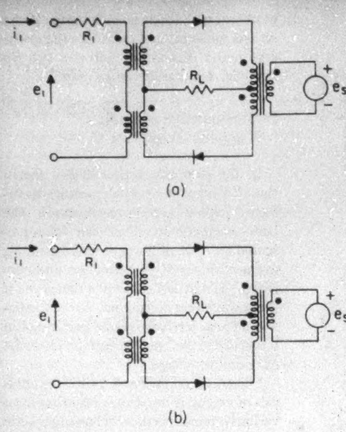


Fig. 14 (left). Diode reversal in magnetic amplifiers

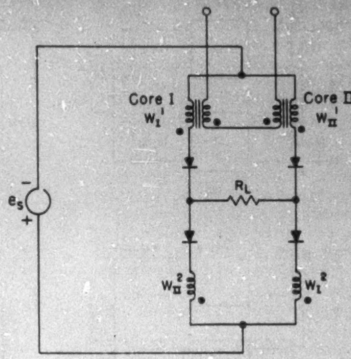


Fig. 15 (right). Bridge circuit

of each basic amplifier, some of the impractical interconnectious may become practical amplifiers. An example of such an amplifier is shown in Fig. 12.

Another group of magnetic amplifiers not considered in the foregoing are the bridge circuits, of which Fig. 15 is an example. Apparently, the bridge circuits cannot be constructed by an interconnection of basic amplifiers. This apparent difficulty can be overcome by splitting the gate winding into two equal parts and by adding another diode in series with the one already in the circuit. The resulting basic magnetic amplifier is shown in Fig. 16. Evidently, the circuit equations of this new basic amplifier are identical to those of the original basic amplifier, since series addition of gate windings and diodes and an interchange in the order of elements connected in series do not effect the operation of the basic amplifier. In Fig. 16 two of these basic amplifiers are connected in series-parallel. The voltage drops between x and x', and y and y' are given by the equations

$$e_{xx}'=e_s-e_x-e_y' \tag{82}$$

$$e_{yy}'=e_s-e_x-e_y' \tag{83}$$

The voltages e_x and e_y, and e_x' and e_y' are equal. Writing Kirchoff's loop equation around the main loop yields

$$2e_s-2e_x-2e_x'=0 \tag{84}$$

or

$$e_x+e_x'=e_s \tag{85}$$

Thus, equations 82 and 83 become

$$e_{xx}'=0$$

$$e_{yy}'=0 \tag{86}$$

The points x and x', and y and y', therefore, can be joined without disturbing the operation of the amplifier; the magnetic-amplifier circuit of Fig. 15 results.

Quite often a resistance is connected in series with the supply source to limit the supply current when both inductors are saturated simultaneously and to improve the amplifier performance. Thus, the voltage across the terminals x and y, and x' and y' depend upon the voltages and currents at the ports. To take into account this effect, the basic amplifier is viewed as a 3-port network, the supply-voltage terminals being the third port. The matrix method of analysis is still applicable since no potential difference exists between the terminals to be joined, as shown by equation 86. In the following subsection the matrix method of determining the behavior of interconnected 2-port networks will be extended to n-port networks.

INTERCONNECTION OF NETWORKS WITH n PORTS

The analysis of interconnected basic amplifiers reduces to simple matrix addition or multiplication if the proper relationships among the port voltages and currents are used. These relations are not so evident for an interconnection of n-port networks as for 2-port networks, because only the Y and Z matrices* exist for the n-port while the $Y,Z,H,G,$ and $ABCD$ matrices exist for the 2-port. The matrix method of analysis, however, can easily be extended to interconnections of networks with an arbitrary number of ports. Any two n-port networks where all the ports are connected pair-wise in series or parallel can be analyzed easily by using the impedance or the admittance matrix relations. It is also evident that a series, or parallel, interconnection of an arbitrary number, m, of n-port networks can be analyzed in the same manner. A series, or parallel, interconnection is obtained when one port from each of the m networks is connected in series, or parallel, with all other $m-1$ like ports and so forth for the remainder of the n ports.

Consider now the case in which ports 1 of all the m networks are connected in series, ports 2 of all the m networks connected in parallel, and so on in a random fashion. The ports can be renumbered so that the port numbers for the series-connected ports run from 1 to p, and from $p+1$ to n for the parallel-connected ports. If the matrix of the interconnected network is to be the sum of the individual network matrices, then all the port quantities that are summed, voltages for series connections and currents for parallel connections, must appear on the left-hand side of the matrix equation or must be chosen as the dependent variables. The port quantities that remain invariant during the interconnections, voltages for parallel connections and currents for series connections, must appear on the right-hand side of the equation or must be chosen as the independent variables. Thus, the proper relationship among the port variables of each network for the interconnection in question is shown by equation 87 where M is an m-by-m matrix.

---

* The scattering matrix and its inverse, the gathering matrix, also exist for an n-port network. To limit the scope of the analysis these matrix relations have been omitted. The usefulness of the scattering matrix in the analysis of magnetic amplifiers have been demonstrated in unpublished notes by G. T. Coate of the Servomechanisms Laboratory at Massachusetts Institute of Technology.

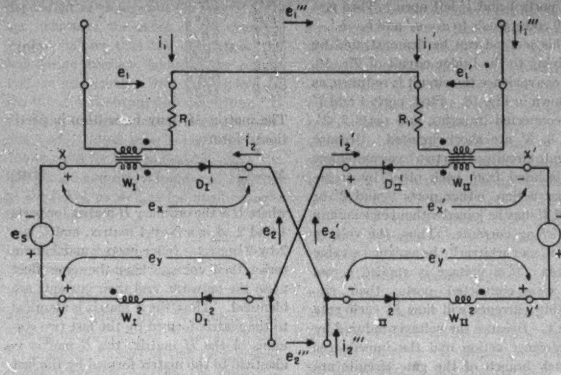


Fig. 16. Series parallel interconnection of two basic amplifiers of the bridge type

$$\left\|\begin{matrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_p \\ i_{p+1} \\ i_{p+2} \\ \cdot \\ \cdot \\ i_n \end{matrix}\right\| = M \left\|\begin{matrix} i \\ i_2 \\ \cdot \\ \cdot \\ i_p \\ e_{p+1} \\ e_{p+2} \\ \cdot \\ \cdot \\ e_n \end{matrix}\right\| \qquad (87)$$

Similar relations can be written for all the other $m-1$ networks. If for network I the following notation is used,

$$\alpha_I = \left\|\begin{matrix} (e_I)_1 \\ \cdot \\ \cdot \\ (e_I)_p \\ (i_I)_{n+1} \\ \cdot \\ \cdot \\ (i_I)_n \end{matrix}\right\| ; \beta_I = \left\|\begin{matrix} (i_I)_1 \\ \cdot \\ \cdot \\ (i_I)_p \\ (e_I)_{n+1} \\ \cdot \\ \cdot \\ (e_I)_n \end{matrix}\right\| \qquad (88)$$

then equation 87 can be written

$$\alpha_I = M_I\beta_I \qquad (89A)$$

Similarly, for the other $m-1$ networks

$$\alpha_{II} = M_{II}\beta_{II}$$

$$\cdot \quad \cdot$$

$$\cdot \quad \cdot$$

$$\alpha_M = M_M\beta_M \qquad (89B)$$

Because ports 1 through p are connected in series and ports $p+1$ through n are connected in parallel, the port voltages and currents of the interconnected network are related to the port voltages and currents of the individual networks by the equations:

$$\left.\begin{matrix} e_k = \sum_{j=I}^{M} (e_j)_k \\ i_k = \sum_{j=I}^{M} (e_j)_k \end{matrix}\right\} \text{for } k=1, 2, \ldots, p \qquad (90)$$

$$\left.\begin{matrix} e_k = (e_j)_k \\ e_k = (i_j)_k \end{matrix}\right\} \begin{matrix}\text{for } j=I, II, \ldots, M \\ \text{and } k=p+1, p+2, \ldots, n\end{matrix} \qquad (91)$$

It is tacitly assumed that the polarities of the port voltages and currents of equations 8(A) and (B) are chosen so that no polarity reversal occurs during the interconnection. Thus, if the matrix equations for the individual networks are added, the matrix equation for the final interconnection is obtained as follows:

$$\alpha = (M_I + M_{II} + \ldots + M_M)\beta \qquad (92)$$

Practical magnetic amplifiers consist of interconnections of identical basic amplifiers. The matrix for these amplifiers, therefore, can be written in the form

$$M = M_I + T_{II}M_IT_{II} + \ldots + T_MM_IT_M \qquad (93)$$

where the T's are the transformation matrices. These matrices account for polarity changes in the port quantities and interchanges of control and gating periods that may occur through interconnections.

For networks constructed from more than two basic elements, or basic amplifiers, there exist other types of interconnections than those considered in the foregoing. A typical example of such an interconnection is the 3-phase amplifier analyzed in reference 1. This amplifier was obtained by connecting two basic elements in series-series and then connecting three of these series-series interconnections in parallel-parallel. A large number of such interconnections exists. Usually, these interconnections can be divided into groups of interconnections, the first group being the interconnection of basic elements to form a number of identical amplifiers. These new amplifiers may then be considered elements for the next group of interconnections, and so forth. The analysis is performed by first determining the relations among port voltages and currents of the first group. Then these relations are used to obtain rela-

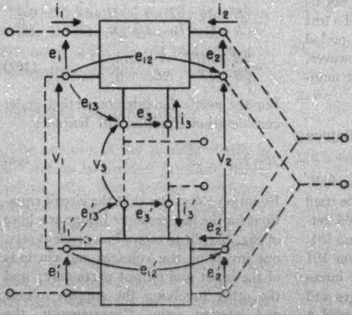


Fig. 17. Validity test for matrix method of determining composite behavior for 3-port networks

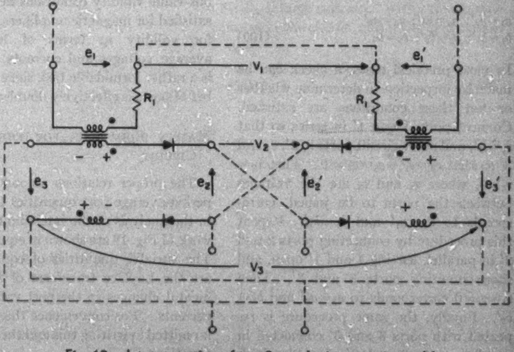


Fig. 18. Interconnection of two 3-port basic magnetic amplifiers

tions among port voltages and currents of the second group, and so forth. Because these groups of amplifiers are formed by different types of interconnections, the relations among port quantities obtained for one group must be rewritten to the form desired for the next group of interconnections in order to apply the matrix method of analysis. This step in the analysis may be involved because matrix inversion of 4-by-4 matrices or larger may be required.

VALIDITY TESTS FOR MATRIX METHOD OF ANALYSIS

The tests that must be performed on a network to insure the validity of the matrix method of determining the composite behavior of interconnected subnetworks are described by Guillemin for ordinary 2-port networks.[2] These same tests can be generalized for the interconnection of n-port networks. For the purpose of illustration, the test procedure for 3-port networks is developed here. Consider the series-parallel-parallel interconnection of Fig. 17, and suppose for the network N that

$$e_{12}=a_1i_1+a_2e_2+a_3e_3 \tag{94}$$

$$e_{13}=b_1i_1+b_2e_2+b_3e_3 \tag{95}$$

and for the network N' that

$$e_{12}'=a_1'i_1'+a_2'e_2'+a_3'e_3' \tag{96}$$

$$e_{13}'=b_1'i_1'+b_2'e_2'+b_3'e_3' \tag{97}$$

For no circulating currents to flow the conditions

$$e_{12}=e_{12}';\ e_{13}=e_{13}' \tag{98}$$

subject to the restrictions that

$$i_1=i_1';\ e_2=e_2';\ e_3=e_3' \tag{99}$$

must be satisfied for all values of i_1, e_2, and e_3. These conditions are satisfied if

$$\begin{aligned}&a_1=a_1';\ a_2=a_2';\ a_3=a_3'\\&b_1=b_1';\ b_2=b_2';\ b_3=b_3\end{aligned} \tag{100}$$

In most practical cases, a check can be made by inspection to determine whether or not these conditions are satisfied. Connect ports 1 and 1′ in series so that $i_1=i_1'$, and short-circuit ports 2, 2′, 3, and 3′ so that $e_2=e_2'=e_3=e_3'=0$. Then $v_2=v_3=0$, where v_2 and v_3 are the voltages between the ports to be joined, corresponds to $a_1=a_1'$ and $b_1=b_1'$. Repeat this procedure by connecting ports 2 and 2′ in parallel, leaving 1 and 1′ open, and short-circuiting ports 3 and 3′. Then $v_1=v_3=0$ corresponds to $a_2=a_2'$ and $b_2=b_2'$. Finally, the same procedure is repeated with ports 3 and 3′ connected in parallel, ports 2 and 2′ short-circuited, and ports 1 and 1′ left open. Then $v_1=v_2=0$ corresponds to $a_3=a_3'$ and $b_3=b_3'$.

This method can be demonstrated by applying to the bridge circuit of Fig. 16. For convenience the circuit is redrawn, as is shown in Fig. 18. First, ports 1 and 1′ are connected in series, and ports 2, 2′, and 3, 3′ are short-circuited. Because the gate circuits of the two basic amplifiers are isolated from each other by transformer action, either ports 2 and 2′ or 3 and 3′ may be joined without causing any circulating currents. Thus, the voltage v_2, or v_3, can arbitrarily be assigned a value of zero. If a voltage is applied across the series-connected ports, then circulating currents will flow in each gate circuit. Because the voltages induced by transformer action and the impedances in each branch of the gate circuits are respectively equal, the voltage v_3 must be zero. In the next step, ports 2 and 2′ are connected in parallel, ports 1 and 1′ are left open, and ports 3 and 3′ are short-circuited. By inspection, v_1 is zero; and from symmetry consideration, v_3 must also be zero. The third test is identical to the one just mentioned since there exists complete symmetry between ports 2 and 2′, and 3 and 3′. Thus, it has been shown that the matrix method of determining the composite behavior is valid for this particular interconnection. It is further evident that any series-parallel-parallel interconnection of basic 3-port bridge elements satisfies the validity tests.

For magnetic amplifiers the validity conditions need only be satisfied for half-period average voltages and currents. If the validity conditions are satisfied at any instant, as in the case of the bridge circuit discussed in the foregoing, then it follows that they also are satisfied for half-period average voltages and currents. It should be pointed out that the instantaneous-value validity conditions need not be satisfied for magnetic amplifiers. To test for validity in terms of half-period average voltages and currents, however, is a rather formidable task since the number of ports is effectively doubled.

MATRIX EQUATIONS FOR THE BRIDGE CIRCUIT

The proper relations among the half-period average port quantities to be used in the analysis of the interconnected network of Fig. 18 are shown in equation 101. The variable quantities of equation 101 are the Laplace transform of the incremental changes in the port voltages and currents. For convenience the symbol Δ is omitted in writing this and the following equations.

$$\begin{Vmatrix}E_1^c\\E_1^g\\I_2^c\\I_2^g\\I_3^c\\I_3^g\end{Vmatrix}=M_I\begin{Vmatrix}I_1^c\\I_1^g\\E_2^c\\E_2^g\\E_3^c\\E_3^g\end{Vmatrix} \tag{101}$$

The matrix M_I may be written in partitioned form

$$M_I=\begin{Vmatrix}H & A\\B & C\end{Vmatrix} \tag{102}$$

where H is the ordinary H matrix for ports 1 and 2, A is a 2-by-4 matrix, and C is a 2-by-2 matrix. Since ports 2 and 3 are in series, their voltages have the same effect upon the network, and their currents are identical. Thus, the A matrix is identical to the matrix formed by the last two columns of the H matrix, the B matrix is identical to the matrix formed by the last two rows, and the C matrix is identical to the h_{22} submatrix.

$$A=\begin{Vmatrix}h_{12}^{cc} & h_{12}^{cg}\\h_{12}^{gc} & h_{12}^{gg}\\h_{22}^{cc} & h_{22}^{cg}\\h_{22}^{gc} & h_{22}^{gg}\end{Vmatrix}=\begin{Vmatrix}h_{12}\\h_{22}\end{Vmatrix} \tag{103}$$

$$B=\begin{Vmatrix}h_{21}^{cc} & h_{21}^{cg} & h_{22}^{cc} & h_{22}^{cg}\\h_{21}^{gc} & h_{21}^{gg} & h_{22}^{gc} & h_{22}^{gg}\end{Vmatrix}=\begin{Vmatrix}h_{21} & h_{22}\end{Vmatrix} \tag{104}$$

$$C=\begin{Vmatrix}h_{22}^{cc} & h_{22}^{cg}\\h_{22}^{gc} & h_{22}^{gg}\end{Vmatrix}=h_{22} \tag{105}$$

The transformation matrix is obtained by inspection and is

$$T'=\begin{Vmatrix}-1 & 0 & 0 & 0 & 0 & 0\\0 & -1 & 0 & 0 & 0 & 0\\0 & 0 & -1 & 0 & 0 & 0\\0 & 0 & 0 & -1 & 0 & 0\\0 & 0 & 0 & 0 & 1 & 0\\0 & 0 & 0 & 0 & 0 & 1\end{Vmatrix}=\begin{Vmatrix}T & 0\\0 & U\end{Vmatrix} \tag{106}$$

where the matrix T is the ordinary 4-by-4 transformation matrix. In general, the matrix of a series-parallel-parallel interconnection of two basic 3-port amplifiers becomes

$$\begin{aligned}M&=M_I+T'M_IT'\\&=\begin{Vmatrix}H & A\\B & C\end{Vmatrix}+\begin{Vmatrix}T & 0\\0 & U\end{Vmatrix}\begin{Vmatrix}H & A\\B & C\end{Vmatrix}\begin{Vmatrix}T & 0\\C & U\end{Vmatrix}\\&=\begin{Vmatrix}H+THT & A+TA\\B+BT & 2C\end{Vmatrix}\end{aligned} \tag{107}$$

For the particular interconnection under consideration equation 107 becomes

$$M=2\begin{Vmatrix}H & 0\\0 & C\end{Vmatrix} \tag{108}$$

Equation 108 shows that the circuit equations for the fast-response bridge amplifier of Fig. 15 split into two independent sets: one involving the voltages and currents of the input and output ports only; and the other involving the supply voltage and current only. Consequently, the half-period average currents and voltages at the input and output ports are independent of changes in the supply voltage, whatever the cause of the changes. It also follows that the changes in the half-period average values of the supply voltage and current are independent of the input and output port quantities. If the ideal waveform restrictions are satisfied and if ideal bias is used, equation 108 is valid for the actual values of the half period average voltages and currents at the ports, as well as for small changes in port quantities.

INTERCONNECTION OF FOUR BASIC MAGNETIC AMPLIFIERS

For the purpose of analysis, a 4-core magnetic amplifier may be viewed as an interconnection of four basic amplifiers. All practical circuits of this type possess complete symmetry so that the number of useful amplifiers is small compared to the number of possible interconnections. No particular new features of interest are introduced by these circuits, and they can be analyzed in the manner already described in the foregoing.

CONCLUSION

The matrix method of analysis discussed in this section is, to the author's knowledge, applicable to all known single-phase magnetic amplifiers. Although this method of analysis may not be the shortest one for a particular circuit, it provides a general means of analysis and gives a clearer insight to the over-all properties of magnetic amplifiers.

Quite often additional control windings are provided for summing purposes. The effect of such additional windings on the amplifier performance can readily be determined by the use of superposition. Feedback is often used to improve the amplifier linearity; its effect upon the amplifier can be analyzed by conventional feedback techniques.

To demonstrate the method of analysis developed, the well-known doubler circuit shown in Fig. 19 will be analyzed in the following section.

Analysis of Doubler Circuit

The analysis of the doubler circuit in this section is divided into two parts: analysis with ideal voltage waveforms, and analysis with nonideal voltage waveforms. Since this analysis is made for the explicit purpose of demonstrating the important ideas developed in this work, the latter part of the analysis is limited to input signal variations for which sequence of transitions *c* occurs. As a result, only part of the amplifier transfer characteristic when resistively loaded is derived.

The doubler circuit may be constructed by connecting two basic amplifiers in series-parallel as shown in Fig. 20. To obtain the proper circuit configuration, it is necessary to reverse the polarity of the supply voltage and the voltage and current at port 2 of basic amplifier *II*. As shown in reference 1 reversal of the supply-voltage polarity may cause the inductors to saturate during alternate half periods. Inspection of the circuit in Fig. 19 indeed shows that the inductors saturate during alternate half periods.

To apply the matrix method of determining the composite behavior, the basic amplifiers must be represented analytically by the H matrix. The H matrix relating incremental signal quantities for the interconnection is of the form

$$H=H_I+TH_{II}T \qquad (109)$$

where H_I and H_{II} are the matrices of the basic amplifiers and T the transformation matrix. The matrix T is obtained by inspection and is

$$T=\begin{Vmatrix}0 & 1 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & 0 & -1\\ 0 & 0 & -1 & 0\end{Vmatrix} \qquad (110)$$

IDEAL VOLTAGE WAVEFORMS

The H matrices of the two basic amplifiers are identical and are given by equation 111

$$H_I=\begin{Vmatrix}R_1'+R_3 & 0 & \dfrac{R_3}{R_b} & 0\\ -R_3\dfrac{1}{z} & R_1' & -\dfrac{R_3}{R_b}\dfrac{1}{z} & 0\\ -\dfrac{R_3}{R_b} & 0 & \dfrac{R_3}{R_mR_b} & 0\\ \dfrac{R_3}{R_f}\dfrac{1}{z} & 0 & \dfrac{R_3}{R_fR_b}\dfrac{1}{z} & \dfrac{1}{R_f}\end{Vmatrix} \qquad (111)$$

where

$$R_1'=\frac{1}{2}R_1 \qquad (112)$$

$$R_3=\frac{R_mR_b}{R_m+R_b} \qquad (113)$$

Substitution of equations 110, 111, and 112 in equation 109 yields

$$H=\begin{Vmatrix}R_1+R_3 & -R_3\dfrac{1}{z} & (-)\dfrac{R_3}{R_b} & \dfrac{R_3}{R_b}\dfrac{1}{z}\\ -R_3\dfrac{1}{z} & R_1+R_3 & \underline{(+)}\dfrac{R_3}{R_b}\dfrac{1}{z} & -\dfrac{R_3}{R_b}\\ \underline{(+)}\dfrac{R_3}{R_b} & \underline{(+)}\dfrac{R_3}{R_b}\dfrac{1}{z} & \dfrac{R_3}{R_mR_b}+\dfrac{1}{R_f} & (-)\dfrac{R_3}{R_fR_b}\dfrac{1}{z}\\ \dfrac{R_3}{R_f}\dfrac{1}{z} & \dfrac{R_3}{R_b} & (-)\dfrac{R_3}{R_fR_b}\dfrac{1}{z} & \dfrac{R_3}{R_mR_b}+\dfrac{1}{R_f}\end{Vmatrix} \qquad (114)$$

Because the output of the doubler circuit is an alternating voltage when the input is a direct voltage, the matrix of equation 114 exhibits the same symmetry property as discussed in reference 1 for the series-connected saturable inductor. Reversal of the polarity of the output voltage and current during every other half period corresponds to pre- and post-multiplying the matrix H by the matrix

$$A=\begin{Vmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & 1\end{Vmatrix} \qquad (115)$$

This operation changes the sign of the matrix elements as indicated in parentheses in equation 114. Using these signs, addition of the first two rows and the last two rows yields

$$E_1=\left\|R_1+R_3\left(1-\frac{1}{z}\right)\right\|I_1-\frac{R_3}{R_b}\left(1-\frac{1}{z}\right)E_2 \qquad (116)$$

$$I_2=\left(\frac{R_3}{R_b}+\frac{R_3}{R_f}\frac{1}{z}\right)I_1+\left\|\frac{1}{R_f}+\frac{R_3}{R_mR_b}-\frac{R_3}{R_fR_b}\frac{1}{z}\right\|E_2 \qquad (117)$$

To obtain the transfer characteristic with resistive load, the following additional equation is needed.

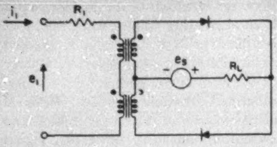
$$E_2=-R_LI_2 \qquad (118)$$

Eliminating I_1 and I_2 from equations 116, 117, and 118 and solving for the ratio E_2/E_1 yield the expression

$$\frac{E_2}{E_1}=-\frac{R_mR_L\left(R_f+R_b\dfrac{1}{z}\right)}{(R_f+R_L)(R_mR_1+R_bR_1+R_mR_b)+(R_1+R_m)R_fR_L-(R_f+R_L)R_mR_b+(R_f+R_1)R_mR_L\dfrac{1}{z}} \qquad (119)$$

If R_m is large compared with other resistances in the circuit, and if the approximation

$$z\cong 1+\frac{\pi}{\omega}s \qquad (120)$$



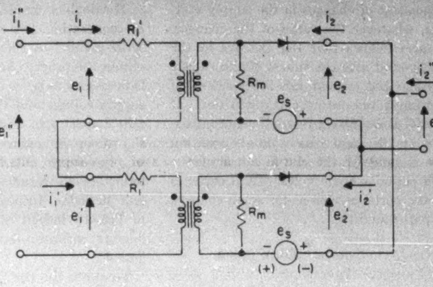


Fig. 19 (left). Doubler circuit

Fig. 20 (right). Doubler circuit as an interconnection of two basic amplifiers

is made, then the following approximate transfer function is obtained.

$$\frac{\bar{E}_2}{\bar{E}_1} \cong -\frac{R_L(R_f+R_b)}{R_fR_1} \times \frac{\dfrac{R_f}{R_f+R_b}\dfrac{\pi}{\omega}s+1}{\left\|\left(1+\dfrac{R_L}{R_f}\right)\left(1+\dfrac{R_b}{R_1}\right)+\dfrac{R_L}{R_1}\right\|\dfrac{\pi}{\omega}s+1} \tag{121}$$

The transfer function of equation 121 can be further simplified by making use of the inequalities: $R_b >> R_f$ and $R_1 >> R_L$ (the latter is not always satisfied).

$$\frac{\bar{E}_2}{\bar{E}_1} \cong -\frac{R_LR_b}{R_fR_1}\frac{1}{\left(1+\dfrac{R_L}{R_f}\right)\left(1+\dfrac{R_b}{R_1}\right)\dfrac{\pi}{\omega}s+1} \tag{122}$$

The transfer function of equation 122 is identical to the one previously derived by the author.[3] A word of caution is necessary in regard to the derivation of equation 122. Equation 122 was obtained, among others, by letting $R_m \to \infty$; this limiting process is not valid if the saturable-inductor loop width is small. In Appendix II it is shown that for the limiting case when $R_m >> R_b$ and the loop width is zero, the diodes conduct during part of the control and block during part of the gating periods of their respective inductors. Bourne and Nitzen have shown that this phenomenon very rarely occurs in practical magnetic amplifiers because of the finite loop-width property of the saturable inductors.[4] In Appendix II it is further shown that the magnetizing resistance alone usually prevents this phenomenon from occurring.

Since the advent of silicon diodes, diode back impedances have become considerably greater than the magnetizing impedance; the inequalities $R_b >> R_f$ and $R_b >> R_m$ then apply so that the transfer function of equation 119 can be written as follows:

$$\frac{\bar{E}_2}{\bar{E}_1} \cong -\frac{R_mR_L}{(R_f+R_L)R_1}\frac{1}{\dfrac{R_1+R_m}{R_1}\dfrac{\pi}{\omega}s+1} \tag{123}$$

Another interesting relation can be obtained from equation 121. Suppose the two diodes in the doubler circuit are short-circuited, then the familiar series-parallel interconnection of two basic elements, basic amplifiers without diodes, results. The approximation that R_m is large compared to the other resistances in the circuit is now valid, and the transfer function of this new amplifier is obtained by substituting $R_f = R_b = R_2$ in equation 121.

$$\frac{\bar{E}_2}{\bar{E}_1} = -\frac{2R_L}{R_1}\frac{\dfrac{\pi}{\omega}s+1}{\left\|\left(1+\dfrac{R_L}{R_2}\right)\left(1+\dfrac{R_2}{R_1}\right)+\dfrac{R_L}{R_1}\right\|\dfrac{\omega}{\pi}s+1} \tag{124}$$

The transfer function derived relates changes in the input and output voltages from their quiescent values. If ideal bias is used, then the steady-state transfer characteristic can be obtained by making the substitution

$z \to 1$

$\bar{E}_1 \to E_1,\ \bar{E}_1 \to I_1,\ I_2 \to I_2$

$\bar{E}_2 \to E_2 + E_s'$

in equations 116 and 117. By making use of the relation $E_2 = -R_LI_2$ and eliminating I_1 and I_2 from the resulting equations, the following expression is obtained:

$$\left\|\frac{1}{R_f}+\frac{1}{R_L}+\frac{R_3}{R_b}\left(\frac{1}{R_m}-\frac{1}{R_f}\right)\right\|E_2 = -\frac{R_3}{R_f}\left(\frac{1}{R_b}+\frac{1}{R_f}\right)E_1 - \left\|\frac{1}{R_f}+\frac{R_3}{R_b}\left(\frac{1}{R_m}-\frac{1}{R_f}\right)\right\|E_s' \tag{125}$$

The voltage E_s' in equation 125 is related to the supply voltage through a transformation similar to the one used in reducing the four by four matrix of equation 114 to a 2-by-2 matrix. This transformation reversed the polarity of the output voltage during the control periods of basic amplifier *1*. From this transformation and because the supply voltage is positive during the control periods and negative during the gating periods, it is seen that the voltage E_s' in equation 125 is always negative. Making use of this property and solving equation 125 for E_2 yields

$$E_2 = \frac{R_L(R_b+R_f)}{R_f(R_m+R_L)+R_b(R_f+R_L)} \times \left\|-\frac{R_m}{R_1}E_1+|E_s|\right\| \tag{126}$$

To compare the results of this and the following subsections, a special case will be considered in which the resistance values and the loop width are related by the equations:

$R_L = R_f = 0.001R$

$R_m = R_b = R$

$R_1 = 0.1R$

$$I_c = \frac{|E_s|}{2R} \tag{127}$$

Substitution of equation 127 in equation 126 gives

$$E_2 = 3.33E_1 + 0.333|E_s| \tag{128}$$

Since $R_L = R_f$, the maximum output is $0.5|E_s|$. The steady-state transfer characteristic is plotted in Fig. 21. The solid line represents the condition of ideal bias, and the dashed line represents the condition of no bias. The two curves are shifted horizontally an amount $R_1I_c = 0.05|E_s|$ from each other.

SINUSOIDAL SUPPLY VOLTAGE AND DIRECT INPUT VOLTAGE

In order to apply the method of analysis developed in the foregoing sections, the voltage waveforms that exist at the input and output ports of each basic amplifier must be known. Because of symmetry these voltage waveforms are the same for both basic amplifiers of which the doubler circuit is constructed. The input ports of the two basic amplifiers are connected in series as shown in Fig. 20. Thus, the voltage that exists across the input ports of each basic amplifier is the sum of the input voltage applied to the doubler circuit and the voltage that exists across the inductor of the other basic amplifier. In Appendix II it is shown that the waveform approximations apply for this case, and that the constant terms in the expressions for the waveform approximations are:

For t near zero,

$$\frac{E_s}{T_s}=\frac{\pi^2}{T}E_s,\ \frac{E_2}{T_2}=0,\ \frac{E_1}{T_1}=\frac{\pi^2}{T'}E_s$$

$$E_1'=E_1 \qquad (129)$$

For t near $T/2$,

$$\frac{E_s}{T_s}=\frac{\pi^2}{T}E_s,\ \frac{E_2}{T_2}=\frac{R_L}{R_L+R_f}\frac{\pi^2}{T}E_s,\ \frac{E_1}{T_1}=0$$

$$E_1''=E_1 \qquad (130)$$

These waveform approximations are valid for sequence of transitions c only. If the direct input voltage is greater than R_1I_c, sequence of transitions b occurs; if it is less than R_1I_c, sequence of transitions c occurs. To limit the analysis to a reasonable length, only sequence of transitions c is considered; and as a result only part of the steady-state transfer characteristic is derived.

The circuit equations for sequence of transitions c are derived in Appendix II. Substituting equations 129 and 130 in equation 162 yields

$$E_R{}^c=c(E_1{}^c-R_1I_c)+d(E_s{}^c+E_2{}^c)+ c\left\|\frac{R_b}{R_1}\frac{R_L+R_f}{R_f}+\frac{R_bR_1}{R_1(R_1+R_b)}\right\|\times \frac{1}{\pi^2E_s}(E_1-R_1I_c)^2 \qquad (131)$$

where E_1 is the input voltage to the doubler circuit. Equation 131, together with equations 44 through 48, determine the steady-state characteristic of the amplifier. Since these equations are nonlinear, the matrix method of determining the composite behavior is not applicable. However, since these equations are valid for any operating point as long as sequence of transitions c occurs, the slope of the steady-state transfer characteristic can be determined by small signal analysis. The matrix method of determining the composite behavior is then applicable since equation 131 can be linearized around any operating point. In terms of incremental signal quantities equation 131 becomes

$$\delta E_R{}^c=c\delta E_1{}^c+d\delta E_2{}^c+cq\delta E_1' \qquad (132)$$

where

$$q=\left\|\frac{R_b}{R_1}\frac{R_L+R_f}{R_f}+\frac{R_bR_1}{R_1(R_1+R_b)}\right\|\times \frac{2}{\pi^2E_s}(E_1-R_1I_c) \qquad (133)$$

and δ is the differential operator.

The term $\delta E_1'$ in equation 132 is the change in the average input voltage to the doubler circuit that occurs during the control periods of basic amplifier I. Thus, $\delta E_1'$ is zero during the gating periods of basic amplifier I. $\delta E_1'$ should not be confused with $\delta E_1{}^c$ which is the change in the average voltage at the input port of amplifier I during the control periods. This latter quantity is the sum of $\delta E_1'$ and the change in the average voltage that appears across the control winding of basic amplifier II during its gating periods. Since the voltages $E_1{}^c$, and $E_1{}^g$, and E_1' can be adjusted independently of each other, these voltages can be considered independent variables.

In terms of increment signal quantities equations 44 through 48 become:

$$\delta E_1{}^c=R_1\delta I_1{}^c+\delta E_R{}^c \qquad (134)$$

$$\delta E_2{}^c=R_b\delta I_2{}^c+\delta E_R{}^c \qquad (135)$$

$$\delta E_1{}^g=R_1\delta I_1{}^g+\delta E_R{}^g \qquad (136)$$

$$\delta E_2{}^g=R_f\delta I_2{}^g+\delta E_R{}^g \qquad (137)$$

$$\delta E_R{}^c+\delta E_R{}^g=0 \qquad (138)$$

From these equations and equation 132 the H matrix of basic amplifier I can be obtained. Since these equations are identical to the equations obtained for the linear case, except for the term $cq\delta E_1'$ in equation 132, the matrix equation of basic amplifier I becomes:

$$\alpha_I=H_I\beta_I+P\delta E_I' \qquad (139)$$

where P is the matrix

$$P=q\left\|\begin{matrix}\frac{c}{1-c}\\ -\frac{c}{1-c}\\ -\frac{c}{1-c}\frac{1}{R_b}\\ \frac{c}{1-c}\frac{1}{R_f}\end{matrix}\right\| \qquad (140)$$

and H_I is obtained by substituting $z=1$ in equation 111. Similarly, for basic amplifier II

$$\alpha_{II}=H_I\beta_{II}+P\delta E_1'' \qquad (141)$$

where $\delta E_1''$ is the change in the half-period average value of the input voltage to the doubler circuit during the control periods of basic amplifier II. Performing the transformation of variables indicated by the equation

$$\alpha_{II}=T\alpha_{II}'$$

$$\beta_{II}=T\beta_{II}' \qquad (142)$$

yields

$$\alpha_{II}'=TH_IT\beta_{II}'+TP\delta E_I'' \qquad (143)$$

To obtain the matrix of the series-parallel interconnected amplifier, equations 139 and 143 are added.

$$\alpha_I+\alpha_{II}'=H_I\beta_I+TH_IT\beta_{II}'+ P\delta E_1'+TP\delta E_1'' \qquad (144)$$

But, since

$$\alpha_I+\alpha_{II}'=\alpha$$

$$\beta_I=\beta_{II}'=\beta \qquad (145)$$

where α and β are the variables of the interconnected amplifier, equation 144 can be rewritten

$$\alpha=\|H_I+TH_IT\|\beta+P\delta E_1'+TP\delta E_1'' \qquad (146)$$

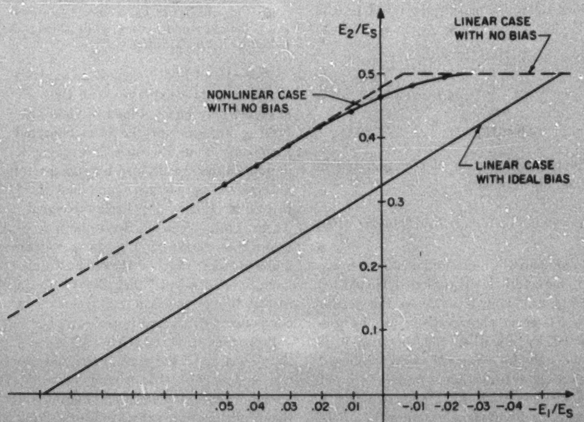


Fig. 21. Steady-state transfer characteristics

The last two terms in equation 146 can be written alternatively:

$$P\delta E_1'+TP\delta E_1''=$$

$$q\left\|\begin{matrix}\frac{c}{1-c} & -\frac{c}{1-c}\\ -\frac{c}{1-c} & \frac{c}{1-c}\\ \underline{(+)}\frac{c}{1-c}\frac{1}{R_b} & \underline{(+)}\frac{c}{1-c}\frac{1}{R_f}\\ \frac{c}{1-c}\frac{1}{R_f} & \frac{c}{1-c}\frac{1}{R_b}\end{matrix}\right\|\left\|\begin{matrix}\delta E_1'\\ \delta E_1''\end{matrix}\right\|$$

$$=qRS \qquad (147)$$

when R and S are the matrices indicated. Ignore, for the time being, the signs in parentheses in equation 147. Again the number of variables can be reduced by making use of the transformation of variables expressed by the matrix of equation 115. Thus, let

$$\alpha=A\alpha' \qquad (148)$$

$$\beta=A\beta'$$

Substitution of equations 147 and 148 in equation 146 yields

$$\alpha'=A\|H_I+TH_IT\|A\beta'+qARS \qquad (149)$$

Premultiplying the matrix R by the matrix A changes the signs in R as indicated in parentheses in equation 147. Addition of the first two rows and the last two rows of the matrix equation, equation 149, yield

$$\delta E_1=R_1\delta I_1 \qquad (150)$$

$$\delta I_2=R_3\left(\frac{1}{R_b}+\frac{1}{R_f}\right)\delta I_1+\left\|\frac{1}{R_f}+\frac{R_3}{R_b}\left(\frac{1}{R_m}-\frac{1}{R_f}\right)\right\|\delta E_2+\frac{2R_3}{R_1}\left(\frac{1}{R_b}+\frac{1}{R_f}\right)q(\delta E_1'+\delta E_1'') \qquad (151)$$

From the definitions of $\delta E_1'$ and $\delta E_1''$, it is seen that

$$\delta E_1'+\delta E_1''=\delta E_1 \qquad (152)$$

To determine the steady-state transfer characteristic with resistive load, equation 152 and the relation $\delta E_2=-R_L\delta I_2$ are substituted in equation 151. Then by using equation 150 to eliminate δI_1, the following differential equation is obtained.

$$\left\|\frac{1}{R_f}+\frac{1}{R_L}+\frac{R_3}{R_b}\left(\frac{1}{R_m}-\frac{1}{R_f}\right)\right\|\delta E_2=-\left(\frac{1}{R_b}+\frac{1}{R_f}\right)\frac{R_3}{R_1}(1+2q)\delta E_1 \qquad (153)$$

The quantity q, as seen from equation 133, is a function of E_1 and may be written in the form

$$q=q'E_1-q'R_1I_c \qquad (154)$$

where

$$q'=\left\|\frac{R_b}{R_1}\frac{R_L+R_f}{R_f}+\frac{R_bR_1}{R_L(R_1+R_b)}\right\|\frac{2}{\pi^2E_s} \qquad (155)$$

A relation between E_1 and E_2 can now be obtained by substituting equation 154 in equation 153 and integrating the resulting equation.

$$E_2=-K\|(1-2q'R_1I_c)E_1+q'E_1^2\|+\text{constant} \qquad (156)$$

where the constant K is the steady-state gain of the amplifier tor the linear case and is given by the expression

$$K=\frac{R_L(R_b+R_f)}{R_f(R_m+R_L)+R_b(R_f+R_L)}\cdot\frac{R_m}{R_1} \qquad (157)$$

The constant of integration may be evaluated in the following manner. When the input signal is equal to R_1I_c, the voltage waveforms are ideal. Consequently, the linear and nonlinear steady-state transfer characteristics should intersect when $E_1=R_1I_c$, provided no bias is used in either case. For the linear case and the chosen parameter values, the output voltage with no bias is given by the dashed line in Fig. 21. As seen from this figure, the output voltage is equal to 0.33 E_s when the input voltage is equal to R_1I_c. For the same parameter values the steady-state transfer characteristic for the nonlinear case becomes

$$E_2=-3.33\left\|0.577E_1+4.235\frac{1}{E_s}E_1^2\right\|+\text{constant}$$

$$=-1.92E_1-14.1\frac{1}{E_s}E_1^2+\text{constant} \qquad (158)$$

Substituting $E_1=R_1I_c=0.05\ E_s$ in equations 158 and equating the result to 0.33 E_s yields

$$\text{constant}=0.333E_s+1.92\times0.05E_s+14.1\frac{1}{E_s}(0.05E_s)^2 \qquad (159)$$

$$=0.464E_s$$

Finally, substitution of equation 159 in equation 158 gives

$$E_2=-1.92E_1-14.1\frac{1}{E_s}E_1^2+0.464E_s \qquad (160)$$

To compare the linear and nonlinear transfer characteristics, equation 160 is plotted in Fig. 21. Only a portion of the steady-state transfer characteristic can be predicted by equation 160, because it is valid only for sequence of transitions c. Sequence of transitions c occurs for input signals less than R_1I_c, and sequence b occurs for input signals greater than R_1I_c. Thus, R_1I_c is the upper limit of the input signal for which equation 160 holds. The waveform approximations are used to calculate the average inductor voltage during the intervals 0 to t_1 and t_2 to $T/2$. These time intervals, therefore, must be small compared to T. When the input voltage is equal to zero, E_1 is approximately $0.05T$ and t_2 is approximately $T/2-0.1T$. Thus, for input signals less than zero, the validity of equation 160 is questionable.

The method of determining the steady-state transfer characteristic demonstrated in the foregoing is not the only one available. For example, the equations describing the steady-state operation of the two basic amplifiers, together with the constraint equations resulting from the interconnection, may be solved directly. Since these equations are nonlinear, considerable algebraic manipulations may be involved in obtaining a solution. On the other hand, the method demonstrated not only determines the steady-state characteristic, but also the small-signal dynamic characteristic.

Summary

The analysis developed in this work is based on five characteristic properties of magnetic amplifiers:

1. Circuits are symmetrical.
2. Circuits consist of interconnections of basic elements.
3. In steady-state operation all voltages and currents are periodic and possess a common period.
4. Each inductor in an amplifier saturates only once per period.
5. For each basic element the part of the period in which the inductor is saturated is less than or equal to the part in which the inductor is unsaturated.

Properties 1 and 2 suggest an analysis in which analytic expressions for the basic elements may be combined to yield analytic expressions for the interconnected amplifier. Thus, the analysis of magnetic amplifiers may be divided into two parts: analysis of basic elements, and analysis of interconnections of basic elements. Three basic elements have been considered, as shown in Fig. 1. From an analysis point of view the first two elements are identical; they differ only in regard to the physical arrangements of components connected in series.

Properties 3, 4, and 5 allow a great simplification of the analytic expressions which determine the steady-state behavior of the basic elements. The derivation of the analytic expressions describing the basic magnetic-amplifier element of reference 1 are, in particular, greatly simplified because the circuit contains only one nonlinear element.

Because the operational cycle of magnetic amplifiers is periodic, a period of operation may be defined. For reasons of analytical simplicity, the period of operation is defined as the time interval between two successive instants in time at which the inductor becomes unsaturated. This period of operation is then divided into two equal parts: the control period and the gating period. Because the output voltage waveform is unrelated to the input voltage waveform, instantaneous values of voltage and current are almost meaningless; and half-period average voltages and currents are chosen as the circuit variables. This adoption of average-quantity variables, along with the defined period of operation, automatically introduces the concept of control- and gating-period average voltages and currents. A control-period average quantity is the average value of the quantity during the control periods; thus, it is a train of pulses where each pulse occurs during a control period and whose magnitude represents the average value that exists during the same period. Similarly, the gating-period average quantities are trains of pulses defined in a like manner for the gating periods. The sum of a control and the corresponding gating priod average voltage or current is thus equal to the actual half-period average voltage or current. The half period average voltages and currents are thus split into components represented by two pulse trains, these pulse trains being displaced from each other in time by $T/2$ sec. This expansion of a signal into a number of pulse trains may be extended by dividing the period of operation into n subintervals, the average values of the current or voltage during the subintervals being the variable quantities. Because the n subintervals repeat periodically with period T, these variable quantities are in the form of n pulse trains, the r^{th} pulse train being displaced in time by T/n sec. from the $(r+1)^{th}$. These n components are orthogonal because each component occupies its own time interval within the period of operation. Friedland employs the same type of expansion in his analysis of periodically time-varying systems, and calls these components "skip-sampled components."[5*]

During the control periods the inductor is unsaturated and its magnetizing impedance can be approximated by a linear resistance if the high permeability, square-loop-type core materials are used. For amplifiers without diodes this approximation is exceedingly good; for amplifiers with diodes the constant d-c loop width must be taken into account. However, if the voltage waveforms satisfy certain conditions, a linear resistance may be used to represent the magnetizing impedance; and, in addition, the diode may be represented as an element which is nonconducting during the control periods and conducting during the gating periods. Thus, during the control periods the basic elements are linear and resistive, and linear relations between control-period average quantities may be written. During the gating periods all components of which the basic elements are composed are linear except for the saturable inductor. Because the operation is periodic the average value of the inductor voltage over a complete period must be zero so that the relation

$$E_R^c+E_R^g=0$$

can be used to eliminate the term E_R^g from the amplifier equations. Thus, the inductor characteristics during the gating periods are immaterial as long as the amplifier possesses properties 3, 4, and 5. The resulting equations are linear relations among control and gating period average voltages and currents at the ports. From these equations, which are based on steady-state properties, the steady-state behavior of magnetic amplifiers can be obtained.

The basic element equations, when recognized as difference equations, can be used to describe the dynamic behavior of the basic elements. The manipulations that are necessary on these equations in order to determine the dynamic behavior are not as simple as for the steady-state case. If the difference equations are transformed through Laplace methods, then all manipulations are reduced to simple algebraic operations, thus simplifying the dynamic analysis. Because the difference equations describing the dynamic properties were derived from steady-state properties, these equations, in general, hold only for small changes in port voltages and currents. However, if certain waveform restrictions are satisfied, these small-change restrictions need not be adhered to; andt he transformed difference equations may be used to characterize the behavior of a magnetic amplifier for large as well as small variations in the port quantities.

To determine the analytic expressions for magnetic amplifiers resulting from interconnections of basic elements, the matrix method of determining the composite behavior, as described by Guillemin for ordinary 2-port networks, may be used.[3] This method is ideally suited for magnetic-amplifier analysis because the input and output ports of basic elements usually are isolated by transformer action. Thus, no circulating current will result from the interconnections of basic elements; this is the validity condition for the matrix method of determining the composite behavior of ordinary linear networks. For magnetic amplifiers resulting from interconnections of basic elements, an additional condition must be satisfied in order to apply the matrix method of determining the composite behavior. The control and gating periods of the individual elements must have the same position in time, or must occur so that the control period of one element corresponds in time to the gating period of the other. The symmetry property of voltage waveforms in magnetic amplifiers usually insures that this condition is satisfied. The extension of the analysis to the polyphase case clearly places in evidence the usefulness of the skip-sampled component expansion, together with the matrix method of determining the composite behavior.

For magnetic amplifiers with diodes the condition that the diodes, or diode, conduct during the gating periods and block during the control periods is not always satisfied. By approximating the voltage waveforms within the regions where this condition is violated, a set of nonlinear relations among the half-period average voltages and currents of the ports can be derived. From these nonlinear relations a set of linear differential equations can be obtained relating differential changes in the average port voltages and currents around a predetermined operating point. Since these differential equations are linear, the matrix method of determining the composite behavior is valid. The resulting differential equations may then be integrated to yield the steady-state characteristics. The small-signal dynamic characteristics are given by the differential equations. This method holds only if the regions in which the voltage-waveform approximations made are small compared to the period of operation.

Appendix I. Circuit Equations for Sequence of Transitions C

In this appendix the basic magnetic-amplifier circuit equations for sequence of transitions c are derived. The objectives

---

[*] The author had already developed the method of analysis presented in this work using the skip-sampled components as variable quantities when he became aware of Friedland's work on periodically time-varying systems.

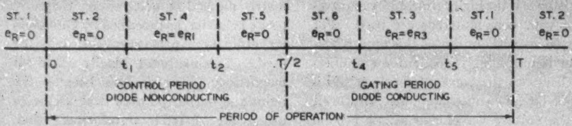


Fig. 22. Sequence of transitions c

of this derivation are twofold: to demonstrate a method by which the equations determining the steady-state or quiescent operating conditions can be obtained, and to determine the voltage waveform restrictions that make these equations linear. Even though this analysis is limited to sequence of transitions c, the method used is equally valid for any one of the other sequences of transitions; sequence of transitions c being chosen because it applies to the doubler circuit of Fig. 19.

For sequence of transitions c, see Fig. 22. Because the diode is nonconducting during the control periods and conducting during the gating periods, equations 44 through 48 are valid. The average inductor voltage during the control period, then, is the only remaining unknown; it is calculated as follows:

$$\begin{aligned}E_R^c&=\frac{2}{T}\int_0^{T/2} e_R dt\\&=\frac{2}{T}\int_0^{T/2}\|(e_1-R_1I_c)+d(e_s+e_2)\|dt-\\&\quad\frac{2}{T}\int_0^{t_1} c\|(e_1-R_1I_c)+d(e_s+e_2)\|dt-\\&\quad\frac{2}{T}\int_{t_2}^{T/2} c\|(e_1-R_1I_c)+d(e_s+e_2)\|dt\end{aligned}\tag{161}$$

where t_1 and t_2 are given by equations 26 and 34, respectively. Substitution of the waveform approximations for e_1, e_s, and e_2 in equation 161 and performing the integration yields

$$\begin{aligned}E_R^c&=c(E_1^c-R_1I_c)+d(E_s^c+E_2^c)+\\&\quad\frac{c}{T\left(\frac{R_1}{R_b}\frac{E_s}{T_s}+\frac{E_1}{T_1}\right)}(E_1'-R_1I_c)^2+\\&\quad\frac{c}{T\left\|\frac{R_1}{R_b}\left(\frac{E_s}{T_s}-\frac{E_2}{T_2}\right)+\frac{E_1}{T_1'}\right\|}\\&\qquad(E_1''-R_1I_c)^2\end{aligned}\tag{162}$$

Equation 162 shows that the half-period average inductor voltage during the control period is a nonlinear function of the slope of the output voltage (E_2/T_2) and the slopes of the input voltage (E_1/T_1 and E_1/T_1') near the zero crossings of the e_1 wave, and of the magnitude of the input voltage at times 0 and $T/2$ (E_1' and E_1'').

The equations for the basic magnetic amplifier are linear if the sum of the last two terms in equation 162 is a constant. For practical amplifiers there are only two cases of interest for which the amplifier equations are linear. These cases are:

$$E_1'=E_1''=R_1I_c \tag{163}$$

$$\left.\begin{aligned}&E_1'=\text{constant}\\&E_1''=\text{constant}\\&\frac{E_1}{T_1}=\text{constant}\\&\frac{E_1}{T_1'}=\text{constant}\\&\frac{E_2}{T_2}=\text{constant}\end{aligned}\right\}\tag{164}$$

If a d-c bias voltage is connected in series with port 1, then equation 163 becomes

$$E_1'=E_1''=R_1I_c-E_b \tag{165}$$

For $E_b=R_1I_c$ (termed ideal bias) the right-hand side of equation 165 is equal to zero. Thus, if the input voltage is zero when the supply voltage is zero and ideal bias is used, then the amplifier equations for sequence of transition c are linear. This result can be generalized for all sequences of transitions. Substitution of equation 163 in equations 20, 26, 28, 34, 35, and 36 yields

$$t_0'=t_1=t_1'=0 \tag{166}$$

$$t_2=t_2'=t_3'=T/2 \tag{167}$$

Thus, the four possible sequences of transitions in Fig. 9 reduce to the following sequence

1–4–6–3–1

The linearity condition of equation 163, therefore, applies for all four sequences of transitions.

The linearization conditions of equation 164 state that both the amplitude and the slope of the input voltage at the zero crossings of the supply voltage wave and the slope of the output voltage wave are constant or independent of the half-period average values of the input and output voltages. The output voltage waveform of resistively loaded magnetic amplifiers invariably possesses these properties. Thus, if the input voltage is obtained from another magnetic amplifier, the linearization conditions of equation 164 are usually satisfied.

Similar results may be obtained by considering any one of the other sequences of transitions. For these sequences a slight complication arises since the diode is both conducting and nonconducting during the control and gating periods. Thus, equations 45 and 47 are not valid. This difficulty can be overcome, for example, by writing equation 45 as follows:

$$E_s^c+E_2^c=R_f(I_2^c)'+R_b(I_2^c)''+E_R^c \tag{168}$$

where $(I_2^c)'$ is the half-period average current at port 2 which flows during the part of the control period in which the diode is conducting, and $(I_2^c)''$ is the half-period average current at port 2 which flows during the part of the control period in which the diode is nonconducting. The total half-period average current at port 2 is

$$I_2^c=(I_2^c)'+(I_2^c)'' \tag{169}$$

Solving equation 169 for $(I_2^c)''$ and substituting the result in equation 168 yields

$$E_s^c+E_2^c=(R_f-R_b)(I_2^c)'+R_bI_2^c+E_R^c \tag{170}$$

Because diode unblocking occurs in the intervals for which the voltage-waveform approximations are valid, the current $(I_2^c)'$ can be expressed in terms of the voltage-waveform approximations. The resulting nonlinear terms can then be combined with the nonlinear terms appearing in the expression for E_R^c.

In steady-state or quiescent operation the port voltages are periodic and of the same period as the supply voltage; equation 162 is thus valid for all periods of operation. From equations 44 through 48 and equation 162 the steady-state or quiescent relations among port voltages and currents can be derived.

If equation 163 is satisfied for all periods of operation, the following operating conditions are met:

1. The operation of the amplifier is periodic and of the same period as the supply voltage.
2. The diode is nonconducting during the control periods and conducting during the gating periods.
3. The inductor voltage is positive during the control periods.
4. The input voltage satisfies the relation

$e_1=R_1I_c$, for $t=T/2$

If these operating conditions are satisfied, then the operation of the amplifier can be described by a set of linear equations among half-period average voltages and currents.

Appendix II. Doubler Circuit

The doubler circuit of Fig. 19 is constructed of basic amplifiers *I* and *II*. In this appendix the sequences of transitions for these basic amplifiers are determined. The conditions chosen are: input of doubler is a direct voltage, supply is a sinusoidal voltage, and a resistive load. The voltage waveform approximations that can be made for sequence of transitions c are then determined, and in the end a discussion of the circuit behavior in the limiting case when $R_m\rightarrow\infty$ is given.

The state of the basic amplifier depends upon the supply and port voltages. Because the doubler circuit is constructed by connecting two basic amplifiers in series-parallel, the input port voltage of basic amplifier *I* is the sum of the direct input voltage to the doubler circuit and the voltage that exists across the control winding of basic amplifier *II*; the output port voltage of each basic amplifier, however, is the same as the output voltage of the doubler circuit.

The currents that flow in the doubler cir-

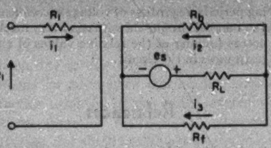


Fig. 23. Equivalent circuit at time zero

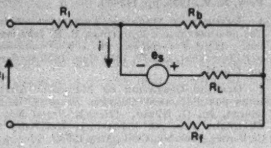


Fig. 24. Approximate equivalent circuit in interval 0 to t_1

cuit, when the supply voltage is zero, are caused by the input voltage only. If the input voltage magnitude is less than R_1I_c, the inductor voltages are zero. A small increase or decrease in the supply voltage, therefore, causes one diode to conduct and the other to block. Thus, the states of the diodes depend entirely upon the supply voltage. A further increase or decrease in the supply voltage causes currents of sufficient magnitude to overcome the loop-width current I_c to flow through the inductors. A voltage now appears across each inductor, driving one towards and the other away from saturation. The events just described correspond to sequence of transitions c. By a similar argument, it can be shown that sequence of transitions b occurs if the input voltage is greater than R_1I_c.

The voltage-waveform approximations for sequence of transitions c will now be determined. The intervals of interest are from time zero to t_1, and from t_2 to $T/2$. An equivalent circuit at time 0+ is shown in Fig. 23. Because R_b is large compared to R_L and R_f, the three loop currents are given by the following equations.

$$i_1=\frac{e_1}{R_1} \tag{171}$$

$$i_2=\frac{e_s}{R_b} \tag{172}$$

$$i_3=\frac{e_s}{R_L+R_f} \tag{173}$$

At time t_1 the net current in inductor I is equal to I_c. The equivalent circuit of Fig. 23, however, is valid only if

$$i_3-i_1<i_1+i_2<I_c \tag{174}$$

Because R_L and R_f are small compared to the other resistances in the circuit, the current i_3-i_1 builds up rapidly and, if $i_1=I_c$, will reach the magnitude I_c at time

$$t_x=\frac{2I_c(R_L+R_L)}{\pi^2E_s}T \tag{175}$$

The time given by equation 175 is the maximum time it takes the current i_3-i_1 to reach I_c. For most practical amplifiers $I_c(R_L+R_f)<0.01\ E_s$ so that

$$t_x<0.002T \tag{176}$$

The time interval from zero to t_x, therefore, is negligible. The time t_1, then, can be determined from the equivalent circuit shown in Fig. 24. The current i can be expressed as follows

$$i=\frac{R_1+R_b}{(R_1+R_f+R_L)R_b}e_s+\frac{1}{R_1+R_L+R_f}e_1 \tag{177}$$

Substituting $e_s=(\pi^2/T)E_st_1$, $e_1=E_1$, and $i=I_c$ in equation 177 and solving for the time t_1 yields

$$t_1=\frac{T}{\pi^2E_s}\frac{R_b}{R_1+R_b}\times \left\|(R_1+R_f+R_L)I_c-\frac{1}{R_1}E_1\right\| \tag{178}$$

If R_1 is of the same order of magnitude as R_f and R_L, the time t_1 is small and may be neglected. Thus, only when R_1 is large compared to R_f and R_L is the time interval from 0 to t_1 of importance; and the approximation that R_1 is large compared to R_f and R_L can be made. Equations 177 and 178 can then be rewritten as follows:

$$i=\frac{1}{R_b}e_s+\frac{1}{R_1}(e_s+e_1) \tag{179}$$

$$t_1=\frac{T}{\pi^2E_s}\frac{R_b}{R_1+R_b}(R_1I_c-E_1) \tag{180}$$

From equations 179 and 180 the voltage-waveform approximations for the time interval from 0 to t_1 become:

$$\frac{E_s}{T_s}=\frac{\pi^2}{T}E_s \tag{181}$$

$$\frac{E_1}{T_1}=\frac{\pi^2}{T}E_s \tag{182}$$

$$E_1'=E_1 \tag{183}$$

An equivalent circuit for the time interval from t_2 to $T/2$ is shown in Fig. 25(A). The waveform approximations then become:

$$\frac{E_s}{T_s}=\frac{\pi^2}{T}E_s \tag{184}$$

$$\frac{E_s}{T_s}=\frac{R_L}{R_L+R_f}\frac{\pi^2}{T}E_s \tag{185}$$

$$\frac{E_1}{T_1'}=0 \tag{186}$$

$$E_1''=E_1 \tag{187}$$

These waveform approximations can be used to determine the average inductor voltage during the control period.

An interesting case arises when R_m is large compared to the other resistances in the circuit. In the analysis that follows ideal waveforms and zero loop width are assumed. These assumptions normally imply that the diodes are nonconducting and conducting respectively during the control and gating periods of their associated inductors. The half-period average input voltage for minimum and maximum output voltage can then be obtained as follows.

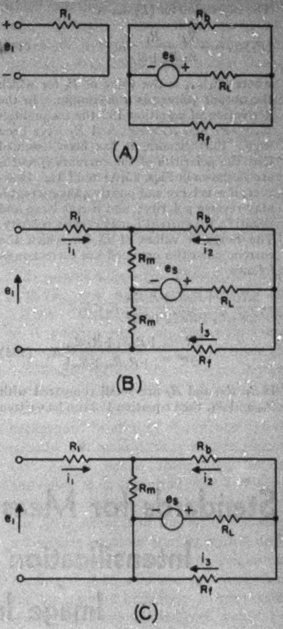


Fig. 25. Equivalent circuit

A—In interval t_2 to T/2
B—When the inductors are unsaturated
C—When one inductor is saturated

Minimum output voltage occurs when neither of the two inductors saturate. This condition is insured if

$$E_{RI}=-E_{RII} \tag{188}$$

The average value of the input voltage for which equation 188 is satisfied can readily be obtained from the equivalent circuit of Fig. 25(B) and is

$$(E_1)_{\min}=\frac{R_1(R_b-R_f)}{(R_f+R_b)(2R_L+R_m+R_b)-R_b(R_b-R_f)}E_s \tag{189}$$

where $(E_1)_{\min}$ is the value of E_1 for which the output voltage is a minimum. Making use of the inequality $R_b>>R_f$ in equation 189 yields

$$(E_1)_{\min}\cong\frac{R_1}{2R_L+R_m}E_s \tag{190}$$

The average value of e_1 for which maximum output voltage is obtained can be determined from the equivalent circuit of Fig. 25(C) and the equation

$$E_{RI}=E_{RII}=0 \tag{191}$$

The expression for $(E_1)_{max}$ is

$$(E_1)_{max} = -\frac{R_f}{R_b}\frac{R_1}{R_L+R_f}E_s \qquad (192)$$

where $(E_1)_{max}$ is the value of E_1 for which the output voltage is a maximum. In the derivation of equation 192 the inequalities $R_m >> R_L$ and $R_f >> R_L$ and R_f have been used. Furthermore, it has been assumed that the polarities of the currents i_2 and i_3 are as shown in Figs. 25(A) and (B). However, if e_1 is large and positive, the current i_2 may reverse polarity; and if e_1 is large and negative, the current i_3 may reverse polarity. The range of values of E_1 for which the current polarities assumed are correct is as follows

$$-\frac{2R_bR_m+R_bR_1+R_1R_m}{2R_LR_m+R_bR_m+R_m^2}E_s < E_1 < \frac{2R_fR_m+R_bR_1+R_LR_m}{2R_LR_m+R_fR_m+R_m^2}E_s \qquad (193)$$

If R_1, R_L, and R_f are small compared with R_m and R_b, then equation 193 can be written

$$-\frac{2R_b}{R_b+R_m}E_s < E_1 < \left\| \frac{2R_f+R_1}{R_m} + \frac{R_bR_1}{R_m^2} \right\| E_s \qquad (194)$$

The useful range of E_1 (range for which the output can be controlled by the input) is obtained from equations 190 and 192, and is

$$-\frac{R_f}{R_b}\frac{R_1}{R_L+R_f}E_s < E_1 < \frac{R_1}{2R_L+R_m}E_s \qquad (195)$$

Because $R_L << R_m$, the right-hand side of equation 194 is always greater than the right-hand side of equation 195. Thus, i_3 is always positive within the useful range of e_1. The left-hand side of equation 193, however, may, if R_m is sufficiently large, be greater than the left-hand side of equation 195. Thus, i_1 may be negative. In the limit when $R_m \to \infty$ both sides of equation 193 are zero, while only the right-hand side of equation 195 is zero. Consequently, if the saturable inductor is ideal, the diodes conduct during the control periods and block during the gating periods, regardless of voltage waveforms. In practical circuits this situation very rarely occurs because of the relative values of the resistances in the circuit.

References

1. Analysis of Magnetic Amplifiers Without Diodes, **Paul R. Johannessen**. *AIEE Transactions*, see pages 471–85 of this issue.

2. Communication Networks, Vol. II (book), **E. A. Guillemin**. John Wiley & Sons, Inc., New York, N. Y., 1935, pp. 145–50.

3. Analysis of Magnetic Amplifiers by the Use of Difference Equations, **P. R. Johannessen**. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 73, 1954 (Jan. 1955 section), pp. 700–11.

4. Dynamic Operation of Magnetic Amplifiers for Feedback Control Systems, **H. C. Bourne, Jr.**, **D. Nitzan**. *Scientific Report No. 1*, Electronics Research Laboratory, University of California, Berkeley, Calif., May 1956.

5. Theory of Time-Varying Sampled-Data Systems, **B. Friedland**. *Technical Report T-191B*, Electronics Research Laboratories, New York, N. Y., Apr. 15, 1957.

Standards for Measurement of Brightness Intensification in Fluoroscopic Image Intensifiers

WALTER S. LUSBY
ASSOCIATE MEMBER AIEE

IT IS convenient to characterize devices for production of intensified X-ray fluoroscopic images in terms of their brightness as compared with conventional fluoroscopic screens. Thus, an X-ray image intensifier is described having an intensification factor of 200 or 500 or 1,000 meaning that the image presented to the viewer is 200 or 500 or 1,000 times brighter than the image would be if he used a conventional X-ray screen. These figures, although widely used today, are meaningless unless all variables are defined and standardized in their measurement. Furthermore, the same term has been used to express the brightness relationship between a light image from a phosphor at the input end of an image device and the light image output. In the case of a fluorescent screen contiguous to a photoelectric surface in a vacuum, the input light image used as the base point can be quite different from, and is usually less bright than, the conventional fluoroscopic screen image under equal stimulus. However, to the best of the author's knowledge, no previous attempt has been made to propose a set of standard conditions on which to base a figure of merit for X-ray image intensifiers.

There are many factors which can lead to variations in the figure of merit for image intensification. First, based on a comparison with a conventional screen, there is the question of how bright the conventional X-ray screen is. The early literature on medical X-ray intensiffers refers to the intensification factor required for success. Obviously the comparison was with a much less bright screen than the conventional screen of today. The industry has progressed from what has been posthumously named the Patterson *A* fluoroscopic screen to the *B*, the *B2*, and the *CB2*, as well as to the screens of other manufacturers, and a steady increase in the brightness of conventional fluoroscopic screens has been witnessed. For example, since the introduction of the Patterson *B* screen in 1934 the average brightness of Patterson fluoroscopic screens has increased by approximately 300%.

Second, the effective wavelength or hardness of the X-ray beam affects the comparisons. Some types of image intensifiers employ a construction such as a heavy glass window in the X-ray beam which acts as a substantial filter. It follows that a comparison of speed of the conventional and the new devices can easily vary by a factor of 2 depending upon the kilovoltage at which the measurements are made.

Third, one can compare brightness by using an open X-ray beam or by using a simulated or actual patient or industrial object. Again a substantial variation in results is obtained.

The need for standards for intensification factors rests not only on technical grounds of scientific correctness, but also on practical grounds. For example, one may attempt to extrapolate cinefluorographic techniques from one type of image intensifier to another using the intensification factors published by the manufacturers and compensating for all other variables. The predicted exposure factors may appear to disagree by more than 2 to 1. In the opinion of the author, this is a direct result of the lack of an accepted common standard and thus there should be unaminous agreement that some form of standard is necessary.

Vol.
Pt. 1
195

Paper 57-316, recommended by the AIEE Electrical Techniques in Medicine and Biology Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted April 26, 1957; made available for printing December 30, 1958.

Walter S. Lusby is with the Westinghouse Electric Corporation, Baltimore, Md.

Another question to be considered is what standard to use. It is not practical to propose a standard which would vary from day to day depending on improvements in phosphors and screen manufacture achieved by the manufacturers of conventional fluoroscopic screens. Rather, one should establish a permanent base line and refer both intensified fluoroscopic images and images from improved conventional screens to this same base line.

For the medical field, the conditions under which the comparisons are made should be simulated clinical conditions. There is considerable latitude for arbitrary choice. The author would like to propose that an X-ray tube voltage of approximately 80 pkv (peak kilovolts) be used for the purpose of the standard. Although techniques tend to shift up or down in kilovoltage over a period of years, 80 pkv appears to be a comfortable midpoint in the range of typical fluoroscopic techniques. For reasons of repeatability of test results a fairly homogeneous X-ray beam is desirable. We use and recommend a 2-cm (centimeters) aluminum filter at the X-ray tube for this purpose. To be more specific and to rule out variations due to generator waveform, length of shockproof cables, etc., it is proposed that this be specified as an X-ray beam with half-value layer of 6.5 millimeters of aluminum obtained with an X-ray tube voltage of 80 pkv±4 pkv.

Another condition to be specified is whether or not a patient is in the beam during measurement, and if so, the patient's thickness. It is proposed that a patient-equivalent phantom be used as close as practical to the image intensifier device, in addition to the filter used at the tube to obtain a homogeneous beam. It is further proposed that the comparisons for the pupose of determining intensification factor be made in three parts representing patient thicknesses near the extremes and center of the absorptions provided by clinical work, and that filters of 2, 4, and 6 cm of 99% pure aluminum be used. These correspond approximately to 10, 20, and 30 cm of boneless solid body tissue.

The author has worked with prestwood phantoms and certainly finds them useful as patient-equivalent filters. Prestwood does vary in density, however, over a fairly substantial range. In addition, the fact that aluminum is less bulky is an advantage.

The X-ray beam should be limited by coning and other precautions should be taken to minimize the effects of scattered radiation. Obviously, the filters which are used as absorbers should extend well beyond the actual cross section of the X-ray beam.

A standard or reference distance from the X-ray source to the fluoroscopic device must be selected. If about 20 cm are allowed between the table top and the screen and the typical 18-inch target-to-table top distance is used, then a target-to-screen distance of 66 cm is selected. It is proposed that this 66 cm be taken as the distance to the outer projection of the device being measured. In other words, assume that a conventional fluoroscopic screen is used with a holder of negligible thickness in contact with the patient at 66 cm from the X-ray tube target. Assume that the device being measured is brought up to this imaginary patient. In some instances, such as in the Westinghouse Fluorex image intensifier, the active input screen will be located more than 66 cm from the X-ray source.

It is proposed that 2 milliamperes of X-ray tube current be accepted as a good practical basis for establishing standard brightness values, and that this be further corrected, if necessary, to provide 94 milliroentgens per minute at 66 cm, measured without backscatter. It is also proposed that the measurements be made at 20 degrees centigrade ±5 degrees and corrected to 20 degrees centigrade.

Because of the extremely low levels of birghtness involved, brightness measurements are difficult, and because the visual mechanism is dual in nature there is an additional problem of color response. At one end of the range of brightness in which we are interested cone vision predominates and color can be seen. At the other end of the range of brightness of interest, rod vision predominates and the eye perceives no color. The use of standard light-measuring instruments such as the Macbeth illuminometer are tedious and are subject to some inaccuracies. It is obvious that in any routine testing one must get away from these basic instruments and provide an instrument which presents a result such as, for example, an electric instrument reading. Still, at one stage of calibration one must rely on light comparisons made by the eye. As a result of many thousands of readings it is proposed that the brightness of conventional fluoroscopy under the proposed conditions be defined at the three specified thicknesses as 8.5, 1.3, and 0.20 microlamberts at 2, 4, and 6 cm of aluminum, respectively.

It is not to be inferred that these numbers, and only these numbers, are satisfactory. A standard would be welcome which uses these or any other logical numbers. Actually, these data represent typical or average screens at the time the data were established several years ago.

With regard to industrial fluoroscopy although its range is somewhat broader than the medical range, the standard as proposed for determining the intensification factor is quite appropriate also for the industrial field.

It is proposed that the same standards be used for all fluoroscopic intensifiers.

To establish the intensification factor of a device one would then proceed as follows. Measure the brightness out-put of the device with the 2-cm aluminum phantom and the standard 80-kv beam at 2 milliamperes and 66-cm target device distance. Turn off the X-ray source and immediately measure the residue or background brightness, if any, and determine by subtraction the brightness due to the X-ray image alone. To avoid ambiguity it is suggested that a residue reading be defined as the value at 1 second after removal of the beam following a 1-minute exposure starting with a dark device. Assume that the brightness reading is 4.25 millilamberts and that the background brightness is zero; then divide the 4.25-millilambert value by 0.0085 and arrive at the intensification factor of 500. The same process is repeated for the 4-cm and 6-cm aluminum phantoms. The three figures are averaged to obtain the standard intensification factor.

One result of adopting this proposed standard would be that all measurements could be made at levels corresponding to the photopic eye and so the photometer used should be designed to have a frequency response corresponding to the response of the photopic eye. The difficulty of measurement of light values in a range near 1 millilambert is significantly less than the difficulty of measurements at the normal fluoroscopic brightnesses. Therefore, the agreement on a standard such as proposed here would provide not only real significance to statements of intensification factor, but also would provide greater accuracy.

References

1. FLUOROSCOPES AND FLUOROSCOPY, W. Edward Chamberlin. *Radiology*, Syracuse, N. Y., vol. 38, 1942, p. 383.

2. FLUOROSCOPIC IMAGE BRIGHTENING BY ELECTRONIC MEANS, John W. Coltman. *Ibid.*, vol. 51, 1948, p. 359.

3. THE INTENSIFICATION OF X-RAY FLUORESCENT IMAGES, W. S. Lusby. *Electrical Engineering*, vol. 70, Apr. 1951, pp. 292–96.

Discussion

J. E. Jacobs (General Electric Company, Milwaukee, Wis.): Mr. Lusby is to be commended for recognizing the situation that exists today regarding the relative intensification factors of commercially available X-ray image intensifiers. While his proposal is a thorough and comprehensive one relating to the image-converter-type amplifier, it does not in my opinion have the basic criterion of a standard, i.e., a measurement related to the fundamental processes involved.

To correct this we propose that the standard, when adopted, relate the image obtainable of a given system under a standard set of conditions to that obtainable under the same conditions by a theoretically perfect amplifier, i.e., one whose image quality is limited only by the statistics of the X-ray image process itself. Computations and studies along these lines are well known and documented.[1-3]

Using this concept of the "ideal amplifier" as the basis for excellence of a particular system, all forms of X-ray image reproduction may be readily compared. This approach has the added advantage of being adaptable to new systems as they are devised.

Unless such a method of standardization is adopted we will be faced with the necessity of devising new standards each time a new system becomes available.

REFERENCES

1. SCREEN INTENSIFICATION SYSTEMS AND THEIR LIMITATIONS, R. E. Sturm, R. H. Morgan. *American Journal of Roentgenology and Radium Therapy*, Springfield, Ill., vol. 62, no. 5, Nov. 1949.
2. ELECTRONIC INTENSIFICATION OF FLUOROSCOPIC IMAGES, M. C. Teves, T. Tol. *Philips Technical Review*, Eindhoven, Netherlands, vol. 14, no. 2, Aug. 1952, pp. 33-68.
3. TECHNICAL SOLUTIONS TO THE PROBLEM OF REDUCING PATIENT DOSAGE, J. E. Jacobs. *AIEE CP 59-236* (available on request).

The Use of Iterated Laplace Transformations in the Solution of Combined Circuit-Field Problems

J. H. MULLIGAN, JR.
MEMBER AIEE

Synopsis: By consideration of a detailed example, the use of iterated Laplace transformations is shown to be an important technique for the solution of combined circuit-field problems, i.e., problems in which the simultaneous solution of ordinary and partial differential equations is necessary. One- and 2-mesh circuits in which eddy currents in a rectangular magnetic core play a significant role are analyzed utilizing the iterated transformations; equivalent circuits characterizing the effect of the eddy currents are also presented.

PROBLEMS sometimes arise in electric circuit theory where both ordinary and partial differential equations must be solved simultaneously in order to provide a rigorous solution to the complete physical problem. In general, the boundary conditions for the partial differential equations will contain unknown quantities which are variables with respect to time in one or more ordinary differential equations associated with the physical system involved. Problems of this kind will be identified as combined circuit-field problems.

A technique that seems particularly suitable for solving problems of the combined circuit-field variety is one using iterated Laplace transformations.[1] By this method, the solution to the partial differential equation(s) may be expressed in terms of the transforms with respect to time of the variables in the ordinary differential equations. This process leads in turn to straightforward simultaneous solution of the (transformed) ordinary differential equations and subsequent direct determination of the desired functions of time.

This paper illustrates the use of the iterated transform technique in the consideration of an example containing 1- and 2-mesh electric circuits which interact with a magnetic core. In addition to the normal circuit equations, the problem requires the solution of a partial differential equation in two space dimensions for a complete description of the electric and magnetic effects involved.

A typical example which involves both field and circuit equations is the eddy current problem which may be formulated with the aid of Fig. 1. (It is of interest to note that although the eddy current problem or one closely related has been treated in detail previously by several investigators, no reference has been found by the writer which treats the problem for a rectangular lamination without the a priori assumption that the problem is one-dimensional rather than the 2-dimensional one which is considered here; see, for example, references 2–9.) A coil of N turns is wound on a re-entrant core of rectangular cross section (area of magnetic material $=A$ square meters) formed from M laminations of width w (meters) and height d (meters). The core, which is of length r (meters), is assumed to be wound uniformly with the N turns. In the range of circuit current values encountered, the magnetic material is assumed to have a constant permeability of μ henrys per meter and a conductivity of σ mhos per meter.

The switch S, which is assumed to have been closed for a sufficient period of time for the current in the coil to have come to equilibrium, is opened at time $t=0$. The current in the coil experiences no discontinuity when the switch is opened; its initial value, assumed known, is designated $i(0^+)$ as indicated in the figure.

Associated with the current in the coil is the storage of energy in the magnetic field produced by the current. When switch S is opened, the stored magnetic energy cannot be dissipated instantly in the absence of a sink capable of absorbing energy at an infinite rate. Consequently, an attempt to change the current in the coil and thus the flux in the magnetic core from their values just prior to opening the

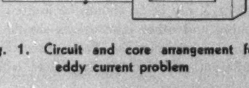


Fig. 1. Circuit and core arrangement for eddy current problem

---

Paper 59-227, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 3, 1958; made available for printing December 3, 1958.

J. H. MULLIGAN, JR., is with New York University, New York, N. Y.

A portion of the work reported herein was done by the author as a consultant to the Bell Telephone Laboratories, Inc., New York, N. Y.

switch results in the production of induced electromotive forces. The latter cause the well-known eddy currents in the magnetic material comprising the core. The decay of current $i(t)$ is thus dependent not only on the nature of the external circuit, in this case composed of the resistor R, but also on the properties of the coil and the core on which it is wound. The potential across the coil can be combined with that across R to obtain a Kirchhoff voltage law equation involving $i(t)$. On the other hand, this potential is a function of the magnetic flux density in the core, a quantity which is related to the current in the coil through a partial differential equation.

Subject to the assumption that conditions in the re-entrant core are well approximated by those in a 2-dimensional model, the flux in the core Φ, the current $i(t)$, and the magnetic field intensity $H(x,y,t)$ are related as follows: (This analysis does not take into account the effects of spin relaxation damping on the decay of flux in the magnetic material. Indications are that whereas this effect is important at lamination thicknesses of 1/4 and 1/8 mil in *4-79* Molybdenum-Permalloy, the classical eddy currents provide the predominant effect in samples of material of approximately 1 mil and greater thickness; see, for example, references 10–11.

$$Ri(t)+N\frac{d\Phi}{dt}=0 \tag{1}$$

where

$$\frac{d\Phi}{dt}=M\frac{d}{dt}\int_0^w\int_0^d B(x,y,t)dydx$$
$$=M\int_0^w\int_0^d \frac{\mu\partial H(x,y,t)}{\partial t}dydx \tag{2}$$

and $H(x,y,t)$ satisfies

$$\frac{\partial^2 H(x,y,t)}{\partial x^2}+\frac{\partial^2 H(x,y,t)}{\partial y^2}=\mu\sigma\frac{\partial H(x,y,t)}{\partial t} \tag{3}$$

plus the initial and boundary conditions

$$H(x,y,0)=\frac{Ni(0^+)}{r} \tag{4}$$

$$H(x,y,t)=\frac{Ni(t)}{r} \tag{5}$$

at $x=0$, w and $y=0$, d.

The quantities M, w, and d are the number, width, and height of the laminations; the co-ordinate system assumed for representation of the laminations is shown in Fig. 2. Conditions 4 and 5 result by considering the line integral of $H(x,y,t)$ around a path enclosing one lamination which, by successive expansion of path, may be shown to be equal to the line integral of $H(x,y,t)$ around the stack of M laminations.

It is convenient to introduce the space average of $H(x,y,t)$ over the cross section of a single lamination. Denoting this quantity by $H_{avg}(t)$, then by definition,

$$wdH_{avg}(t)=\int_0^w\int_0^d H(x,y,t)dydx \tag{6}$$

and equation 1 can now be written

$$Ri(t)+\mu AN\frac{dH_{avg}(t)}{dt}=0 \tag{7}$$

where

A = total cross-sectional area of magnetic material in the core

To determine $i(t)$, equations 3 and 7 must be solved simultaneously subject to conditions 4 and 5. To effect the solution, it is convenient to utilize iterated Laplace transforms. First, the transform of the two equations with respect to time is taken which converts equation 7 to an algebraic equation and leaves equation 3 containing only derivatives with respect to x and y. Laplace transformation of the transformed equation 3 with respect to x then yields an ordinary differential equation in y. Solution of this equation and subsequent inversion of the transform equation yields a transform solution for H (with only time transformed) that may be solved simultaneously with the transform of equation 7 to find $i(t)$.

Formulation and Solution of Transform Equations

Laplace Transformation With Respect to Time

Let the Laplace transforms of $H_{avg}(t)$ and $i(t)$ be denoted as $\bar{H}_{avg}(s)$ and $I(s)$ where

$$\bar{H}_{avg}(s)=\mathfrak{L}[H_{avg}(t)]=\int_0^\infty H_{avg}(t)\epsilon^{-st}dt$$

and

$$I(s)=\mathfrak{L}[i(t)]=\int_0^\infty i(t)\epsilon^{-st}dt$$

Taking the Laplace transform of equation 7 then produces

$$RI(s)+\mu AN[s\bar{H}_{avg}(s)-H_{avg}(0^+)]=0 \tag{8}$$

Utilizing condition 4 and the defining equation for self-inductance $L=\mu N^2A/r$, equation 8 can be written in the form

$$RI(s)+\mu ANs\bar{H}_{avg}(s)=Li(0^+) \tag{9}$$

Once $\bar{H}_{avg}(s)$ can be expressed as a function of $I(s)$, equation 9 will permit solution for $I(s)$, which upon inversion, yields $i(t)$. Since $i(t)$ is involved in the boundary conditions for the partial differential equation 3, the desired expression for $\bar{H}_{avg}(s)$ can be obtained by solving that equation in transform form.

Taking the Laplace transform of equation 3 and the initial and boundary conditions 4 and 5 yields

$$\frac{\partial^2\bar{H}(x,y,s)}{\partial x^2}+\frac{\partial^2\bar{H}(x,y,s)}{\partial y^2}=\mu\sigma\bar{H}(x,y,s)-\mu\sigma\frac{Ni(0^+)}{r} \tag{10}$$

and

$$\bar{H}(x,0,s)=\bar{H}(x,d,s)=\bar{H}(0,y,s)=\bar{H}(w,y,s)=\frac{NI(s)}{r} \tag{11}$$

where

$$\bar{H}(x,y,s)=\mathfrak{L}_t[H(x,y,t)]=\int_0^\infty H(x,y,t)\epsilon^{-st}dt$$

Solution of Equation for $\bar{H}(x,y,s)$

The Laplace transform of the function $\bar{H}(x,y,s)$ with respect to the variable x is defined as

$$\mathfrak{L}_x[\bar{H}(x,y,s)]=H^*(z,y,s)=\int_0^\infty \bar{H}(x,y,s)\epsilon^{-zx}dx$$

Transforming equations 10 and 11 results in the expressions

$$\frac{d^2H^*(z,y,s)}{dy^2}+(z^2-\mu\sigma s)H^*(z,y,s)=z\frac{NI(s)}{r}-\mu\sigma\frac{Ni(0^+)}{r}+\varphi(y,s) \tag{12}$$

$$H^*(z,0,s)=H^*(z,d,s)=\frac{NI(s)}{rz} \tag{13}$$

In equation 12 the partial derivative with respect to y has been replaced by an ordinary derivative and $\varphi(y,s)$ has been used to designate the function, as yet unknown,

$$\frac{\partial\bar{H}(x,y,s)}{\partial x}$$

evaluated at $x=0$.

The solution of equation 12 subject to the boundary conditions (equation 13) is[1,12]

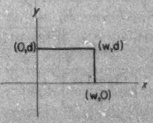


Fig. 2. Geometry of core cross section

$$H^*(z,y,s)=\frac{zNI(s)}{r(z^2-\mu\sigma s)}-\frac{\mu\sigma Ni(0^+)}{rz(z^2-\mu\sigma s)}+$$
$$\frac{1}{\sin d\sqrt{z^2-\mu\sigma s}}\left\{\left[-\frac{NI(s)z}{r(z^2-\mu\sigma s)}+\frac{\mu\sigma Ni(0^+)}{rz(z^2-\mu\sigma s)}+\frac{NI(s)}{rz}\right]\times\right.$$
$$\left.\left[\sin y\sqrt{z^2-\mu\sigma s}+\sin(d-y)\sqrt{z^2-\mu\sigma s}\right]\right\}-$$
$$\frac{1}{\sin d\sqrt{z^2-\mu\sigma s}}\left\{\frac{\sin y\sqrt{z^2-\mu\sigma s}}{\sqrt{z^2-\mu\sigma s}}\int_y^d\varphi(\eta,s)\sin(d-\eta)\sqrt{z^2-\mu\sigma s}\,d\eta+\right.$$
$$\left.\frac{\sin(d-y)\sqrt{z^2-\mu\sigma s}}{\sqrt{z^2-\mu\sigma s}}\int_0^y\varphi(\eta,s)\sin\eta\sqrt{z^2-\mu\sigma s}\,d\eta\right\}\qquad(14)$$

The function $\bar{H}(x,y,s)$ is to be determined from $H^*(z,y,s)$ by the inverse Laplace transformation, namely

$$\bar{H}(x,y,s)=\mathfrak{L}_z^{-1}[H^*(z,y,s)]$$
$$=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}H^*(z,y,s)\epsilon^{zx}dz\qquad(15)$$

This integral can be evaluated utilizing contour integral techniques, but to do so, the nature and location of the singularities of the integrand must be known. Detailed examination of the terms that comprise $H^*(z,y,s)$ indicates that only simple poles are present. The pole locations and the residues of $\epsilon^{zx}H^*(z,y,s)$ in these poles will be considered on a term by term basis.

Examination of the behavior of the first, second, and third terms in the expression for $H^*(z,y,s)$ indicates that a removable singularity exists at $z=\pm\sqrt{\mu\sigma s}$. No other singularity results from the first term, although the pole at $z=0$ in the second term contributes an amount $Ni(0^+)/sr$ as the residue of $\epsilon^{zx}H^*(z,y,s)$ in this pole. The corresponding contribution of the pole at $z=0$ in the third term is

$$\left[\frac{NI(s)}{r}-\frac{Ni(0^+)}{rs}\right]\times$$
$$\left[\frac{\sinh y\sqrt{\mu\sigma s}+\sinh(d-y)\sqrt{\mu\sigma s}}{\sinh d\sqrt{\mu\sigma s}}\right]$$

The third term also has simple poles at

$$z=\pm\sqrt{\frac{(2n-1)^2\pi^2}{d^2}+\mu\sigma s}\ (n=1,2,3,\ldots)$$
$$\text{for } y\neq 0,\ y\neq d$$

which contributes the following amount to the evaluation of the integral:

$$4\frac{\mu\sigma s}{(2n-1)\pi}\left[\frac{NI(s)}{r}-\frac{Ni(0^+)}{sr}\right]\times$$
$$\frac{\sin\frac{(2n-1)\pi y}{d}\cosh x\sqrt{\left[\frac{(2n-1)\pi}{d}\right]^2+\mu\sigma s}}{\left[\frac{(2n-1)\pi}{d}\right]^2+\mu\sigma s}$$

The fourth and fifth terms have simple poles at

$$z=\pm\sqrt{\left(\frac{n\pi}{d}\right)^2+\mu\sigma s}\ (n=1,2,3,\ldots)$$
$$\text{for } y\neq 0,\ y\neq d$$

resulting in the contribution of the following term:

$$\sin\frac{n\pi y}{d}\sinh x\sqrt{\left(\frac{n\pi}{d}\right)^2+\mu\sigma s}\times$$
$$\frac{2}{d}\frac{\int_0^d\varphi(\eta,s)\sin\frac{n\pi\eta}{d}d\eta}{\sqrt{\left(\frac{n\pi}{d}\right)^2+\mu\sigma s}}$$

With a knowledge of the poles of $H^*(z,y,s)$ and the residues of $\epsilon^{zx}H^*(z,y,s)$ in the poles, as well as the realization that $H^*(z,y,s)$ vanishes as z becomes infinite, the theorem of residues may be applied directly to the evaluation of equation 15. Combining the residues just obtained, it is found that

$$H(x,y,s)=\frac{Ni(0^+)}{rs}+\left[\frac{NI(s)}{r}-\frac{Ni(0^+)}{rs}\right]\left[\frac{\sinh y\sqrt{\mu\sigma s}+\sinh(d-y)\sqrt{\mu\sigma s}}{\sinh d\sqrt{\mu\sigma s}}\right]+$$
$$\mu\sigma s\left[\frac{NI(s)}{r}-\frac{Ni(0^+)}{rs}\right]\frac{4}{\pi}\sum_{n=1}^{\infty}\frac{1}{2n-1}\frac{1}{\left[\frac{(2n-1)\pi}{d}\right]^2+\mu\sigma s}\times$$
$$\cosh x\sqrt{\left[\frac{(2n-1)\pi}{d}\right]^2+\mu\sigma s}\ \sin\frac{(2n-1)\pi y}{d}+$$
$$\sum_{n=1}^{\infty}\frac{g_{n(s)}}{\sqrt{\left(\frac{n\pi}{d}\right)^2+\mu\sigma s}}\sinh x\sqrt{\left(\frac{n\pi}{d}\right)^2+\mu\sigma s}\ \sin\frac{n\pi y}{d}\qquad(16)$$

where

$$g_n(s)=\frac{2}{d}\int_0^d\varphi(\eta,s)\sin\frac{n\pi\eta}{d}d\eta\qquad(17)$$

and $y\neq 0$, $y\neq d$.

When $y=0$ or $y=d$, the poles from which the third and fourth terms were derived vanish, so that $\bar{H}(x,y,s)$ reduces to the one term $NI(s)/r$ as it should. The function $\varphi(\eta,s)$ may be replaced by a known quantity by the use of one of the boundary conditions 11, namely $\bar{H}(w,y,s)=NI(s)/r$. For the evaluation, the second term of equation 16 is expanded in a Fourier sine series and then with x set equal to w, expression 16 is set equal to $NI(s)/r$ in view of the boundary condition cited. This results in the following equality:

$$1=\frac{4}{\pi}\sum_{n=1}^{\infty}\frac{1}{2n-1}\times$$
$$\left[\frac{(2n-1)\pi}{d}\right]^2+\mu\sigma s\cosh w$$
$$\frac{\sqrt{\left[\frac{(2n-1)\pi}{d}\right]^2+\mu\sigma s}}{\left[\frac{(2n-1)\pi}{d}\right]^2+\mu\sigma s}\times$$
$$\sin\frac{(2n-1)\pi y}{d}+\frac{4}{\pi}\sum_{n=1}^{\infty}a_n$$
$$\sinh w\sqrt{\left(\frac{n\pi}{d}\right)^2+\mu\sigma s}\ \sin\frac{n\pi y}{d}$$

where

$$a_n=\frac{\pi}{4}\frac{g_n(s)}{\left[\frac{NI(s)}{r}-\frac{Ni(0^+)}{rs}\right]\sqrt{\left(\frac{n\pi}{d}\right)^2+\mu\sigma s}}$$

Replacing the factor unity on the left-hand side of the equation by the Fourier sine series,

$$1=\frac{4}{\pi}\sum_{n=1}^{\infty}\frac{1}{2n-1}\sin\frac{(2n-1)\pi y}{d},\ 0<y<d$$

and equating coefficients, produces the final result

$$a_n=0\ (n\text{-even})$$

$$a_n=\frac{\mu\sigma s}{n}\times$$
$$\frac{\left[1-\cosh w\sqrt{\left(\frac{n\pi}{d}\right)^2+\mu s}\right]}{\left[\left(\frac{n\pi}{d}\right)^2+\mu\sigma s\right]\sinh w\sqrt{\left(\frac{n\pi}{d}\right)^2+\mu\sigma s}}$$
$$(n\text{-odd})$$

The substitution of the last-named result in equation 16 allows the equation for $H(x,y,s)$ to be simplified to the form

$$H(x,y,s)=\frac{Ni(0^+)}{rs}+\left[\frac{NI(s)}{r}-\frac{Ni(0^+)}{rs}\right]\frac{4}{\pi}\times$$

$$\sum_{n=1}^{\infty}\frac{\left[\frac{(2n-1)\pi}{d}\right]^2+\mu\sigma s\left[\frac{\sinh x\sqrt{\left[\frac{(2n-1)\pi}{d}\right]^2+\mu\sigma s}+\sinh(w-x)\sqrt{\left[\frac{(2n-1)\pi}{d}\right]^2+\mu\sigma s}}{\sinh w\sqrt{\left[\frac{(2n-1)\pi}{d}\right]^2+\mu\sigma s}}\right]}{(2n-1)\left[\frac{(2n-1)^2\pi^2}{d^2}+\mu\sigma s\right]}\times$$

$$\sin\frac{(2n-1)\pi y}{d} \qquad (18)$$

Equation 18 is of considerable interest inasmuch as it expresses the transform of $H(x,y,t)$ in terms of the known initial value of $i(t)$ and the transform of $i(t)$. If, for example, the current in the coil wound on the core is forced to have a certain known time variation, equation 18 allows the direct determination of the magnetic field as a function of time at all points in the cross section of the core. On the other hand, formulation of $H(x,y,s)$ in terms of $I(s)$ as in equation 18 permits determination of $I(s)$ as well as the magnetic field in a problem such as the one under consideration where equations must be solved simultaneously to obtain the two functions.

Before proceeding to the calculation of $\bar{H}_{avg}(s)$ and the solution for $I(s)$, it is of interest to examine the variation of $\bar{H}(x,y,s)$ in various limiting situations. At the boundaries $x=0$ and $x=w$, it is clear that the numerator cancels all but the factor n in the denominator of the second term. The series that remain sums to unity with the result that $\bar{H}$ goes to $NI(s)/r$ as it should. For $y=0$ and $y=d$, it has already been mentioned in connection with equation 16 that $\bar{H}(x,y,s)=NI(s)/r$. Finally, the behavior of $H(x,y,t)$ as t approaches zero or becomes infinite can be found from equation 18 by the limiting processes

$$H(x,y,0)=\lim_{s\to\infty}s\bar{H}(x,y,s)$$

$$H(x,y,\infty)=\lim_{s\to 0}s\bar{H}(x,y,s)$$

In connection with the first limiting process, it is seen from equation 18 that the factor containing the hyperbolic sines approaches zero ($x\neq 0\neq w$) as s becomes infinite with the result that

$$H(x,y,0)=\lim_{s\to\infty}s\bar{H}(x,y,s)=\frac{Ni(0^+)}{r}$$

Similarly, as s approaches zero, the second term in equation 18 approaches the Fourier sine series for unity with the result

$$H(x,y,\infty)=\lim_{s\to 0}s\bar{H}(x,y,s)$$

$$=\lim_{s\to 0}\left[\frac{Ni(0^+)}{r}+\frac{NsI(s)}{r}-\frac{Ni(0^+)}{r}\right]$$

$$=\frac{Ni(\infty)}{r}$$

These results are in accordance with the stated initial condition and the physical nature of the final condition for the equation.

DETERMINATION OF $I(s)$

The average value of $\bar{H}(x,y,s)$ over the cross section of the rectangular region bounded by the lines $x=0$, $x=w$ and $y=0$, $y=d$ is defined as

$$wd\bar{H}_{avg}(s)=\int_0^w\int_0^d\bar{H}(x,y,s)dydx$$

which, when applied to equation 18, yields

$$\bar{H}_{avg}(s)=\frac{Ni(0^+)}{rs}+\left[\frac{NI(s)}{r}-\frac{Ni(0^+)}{rs}\right]\frac{8}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{(2n-1)^2}\times$$

$$\frac{\frac{(2n-1)^2\pi^2}{d^2}+\mu\sigma s\frac{\tanh\frac{w}{2}\sqrt{\left[\frac{(2n-1)\pi}{d}\right]^2+\mu\sigma s}}{\frac{w}{2}\sqrt{\left[\frac{(2n-1)\pi}{d}\right]^2+\mu\sigma s}}}{\left[\frac{(2n-1)\pi}{d}\right]^2+\mu\sigma s} \qquad (19)$$

An examination of the second term in equation 19 reveals that the term in the expansion for each value of n is a function which has only poles on the negative real axis with zeros on that axis which separate the poles. The poles are located at

$$s=-\rho_{mn}=-\frac{\pi^2}{\mu\sigma}\left[\frac{(2n-1)^2}{d^2}+\frac{(2m-1)^2}{w^2}\right];$$

$$m,n=1,2,3,\ldots \qquad (20)$$

with residues (for the function inside the summation sign)

$$k_{mn}=\frac{8}{\mu\sigma}\left[\frac{1}{(2n-1)^2w^2}+\frac{1}{(2m-1)^2d^2}\right] \qquad (21)$$

Defining the function $F(s)$ as

$$F(s)=\frac{64}{\pi^2\mu\sigma}\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\times$$

$$\frac{\frac{1}{(2n-1)^2w^2}+\frac{1}{(2m-1)^2d^2}}{s+\frac{\pi^2}{\mu\sigma}\left[\frac{(2n-1)^2}{d^2}+\frac{(2m-1)^2}{w^2}\right]} \qquad (22)$$

the expression for $H_{avg}(s)$ may be written in the form

$$H_{avg}(s)=\frac{Ni(0^+)}{rs}+\left[\frac{NI(s)}{r}-\frac{Ni(0^+)}{rs}\right]F(s) \qquad (23)$$

Equations 9 and 23 may then be combined to furnish the following single equation for $I(s)$

$$RI(s)+sLF(s)I(s)=Li(0^+)F(s) \qquad (24)$$

whence

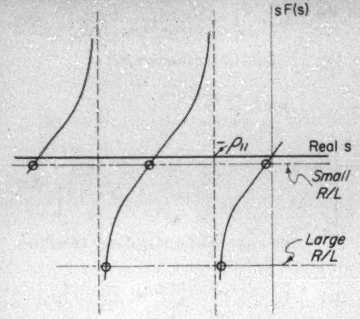
$$I(s)=\frac{Li(0^+)F(s)}{R+sLF(s)}=\frac{\frac{L}{R}i(0^+)}{\frac{sL}{R}+\frac{1}{F(s)}} \qquad (25)$$

Solution for i(t)

The expression, equation 25, for $I(s)$ has simple poles at those values of s for which

$$sF(s)=-\frac{R}{L} \qquad (26)$$

The dependence of the form of $i(t)$ both on the circuit external to the coil as well as on the characteristics of the core laminations can be demonstrated from equation 26. The inductance parameter L is dependent on the number of turns in the coil as well as the permeability, cross-sectional area, and length of the magnetic core, but it is not affected by either the conductivity of the magnetic material or the dimensions of the individual laminations. The quantity $F(s)$, on the other hand, is a function only of the dimensions of the individual laminations and the



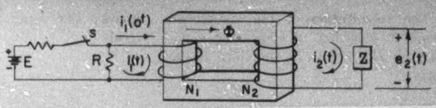


Fig. 3 (left). The variation of the roots of equation 26 with R/L ratio

Fig. 4 (above). Circuit and core arrangement for 2-mesh problem

permeability and conductivity of the magnetic material. Two extreme situations may be compared to illustrate the different effects produced on $i(t)$ by the combined effect of the circuit parameters and lamination characteristics.

$F(s)$ is a function having only poles and zeros which separate one another on the negative real axis. $F(0)$ has the value unity. The first pole is located at

$$s=-\rho_{11}=-\frac{\pi^2}{\mu\sigma}\left(\frac{1}{d^2}+\frac{1}{w^2}\right)$$

and for value of s on the negative real axis whose magnitudes are appreciably less than ρ_{11}, $F(s)$ is approximately unity. As a consequence, if R/L is small compared to ρ_{11}, the first root of equation 26 is approximately $-R/L$. Fig. 3 depicts the behavior of the function $sF(s)$ and indicates the location of the roots of equation 26 for two extreme values of the ratio R/L. The roots of the equation, shown by circles, occur at the intersections of the horizontal line displaced $-R/L$ units from the real axis with the curve representing $sF(s)$. For small values of the ratio, the poles of $I(s)$ (see equation 25) are near the zeros of $sF(s)$ whereas for large values of R/L, the poles of $I(s)$ are located in the vicinity of the poles of $sF(s)$.

An estimate of the roots of equation 26 may be made under the condition of a finite but very large value of R resulting in a high R/L ratio. In view of the Laurent expansion for $F(s)$ in the vicinity of the pole at $s=-\rho_{mn}$, it is permissible to write

$$F(s)\doteq\frac{k_{mn}}{s+\rho_{mn}} \tag{27}$$

when s is approximately equal to $-\rho_{mn}$ where k_{mn} and ρ_{mn} have the values given by equations 20 and 21. Insertion of approximation 27 in the root equation 26 yields the following approximate values for the locations of the poles of $I(s)$:

$$s=-\frac{\rho_{mn}}{1+\dfrac{k_{mn}L}{R}} \tag{28}$$

Evaluation of the residues in these poles allows the transform of $I(s)$ to be expressed as

$$I(s)\doteq Li(0^+)\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{k_{mn}\left(1+\dfrac{k_{mn}L}{R}\right)^{-1}}{s+\dfrac{\rho_{mn}}{1+\dfrac{k_{mn}L}{R}}} \tag{29}$$

In the event that R/L is not large enough for approximation 28 to be valid, $I(s)$ can be written in the general form

$$I(s)=Li(0^+)\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{k_{mn}'}{s+\rho_{mn}'} \tag{30}$$

where the ρ_{mn}' are the roots of equation 26 and k_{mn}' are the corresponding residues.

Inversion of the expression for $I(s)$ yields

$$i(t)=Li(0^+)\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}k_{mn}'\epsilon^{-\rho'_{mn}t} \tag{31}$$

for the situation where R/L has an arbitrary value and

$$i(t)\doteq Li(0^+)\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{k_{mn}\epsilon^{-\rho_{mn}t}}{1+\dfrac{k_{mn}L}{R}} \tag{32}$$

where R/L is very large. The largest value of k_{mn} occurs when $m=n=1$ and the largest value of the ratio $k_{mn}L/R$ is, therefore,

$$\frac{k_{11}L}{R}=\frac{\dfrac{64}{\pi^4}\dfrac{\pi}{\mu\sigma}\left[\dfrac{1}{w^2}+\dfrac{1}{d^2}\right]}{R/L}=\frac{64}{\pi^4}\frac{\rho_{11}}{R/L}$$

Clearly, the expression "large values of R/L" is to be interpreted as R/L being large compared to the first pole in the representation of the complex impedance function $sLF(s)$ resulting from the presence of eddy currents.

Equation 31 together with Fig. 3, which depicts the dependence of the roots ρ_{mn}' on the ratio R/L, indicates how the current $i(t)$ depends both on the circuit parameters R and L as well as the lamination characteristics. An example in which the joint dependence is more involved is considered in the next section.

Extension to 2-Mesh Problem

Fig. 4 indicates a modification of the system of Fig 1 which has just been analyzed. A coil of N_2 turns has been wound uniformly on the core and connected to the arbitrary impedance Z. It is desired to find expressions for $I_1(s)$, $I_2(s)$, and $E_2(s)$, the Laplace transforms of $i_1(t)$, $i_2(t)$, and $e_2(t)$ which are shown in the figure.

Representing the integrodifferential operator of the impedance Z by $Z(p)$ with Laplace transform $Z(s)$ (in the absence of energy storage), the equilibrium equations for the system can be written

$$Ri_1(t)+\mu AN_1\frac{dH_{\text{avg}}}{dt}=0$$

$$Z(p)i_2(t)+\mu AN_2\frac{dH_{\text{avg}}}{dt}=0 \tag{33}$$

which after transformation become

$$RI_1(s)+\mu N_1A\left[s\bar{H}_{\text{avg}}(s)-H_{\text{avg}}(0^+)\right]=0$$

$$Z(s)I_2(s)+\mu N_2A\left[s\bar{H}_{\text{avg}}(s)-H_{\text{avg}}(0^+)\right]=0 \tag{34}$$

where it is assumed there is no initial energy storage in the impedance Z. If energy is stored in Z at $t=0$, terms appropriate to the modes of energy storage must be added to the second equation 34. These two equations must be solved simultaneously with the partial differential equation 3 subject, however, to the following initial and boundary conditions:

$$H(x,y,0^+)=\frac{N_1i_1(0^+)+N_2i_2(0^+)}{r} \tag{35}$$

$$H(x,y,t)=\frac{N_1i_1(t)+N_2i_2(t)}{r}\text{ at }x=0,\ w;\quad y=0,\ d \tag{36}$$

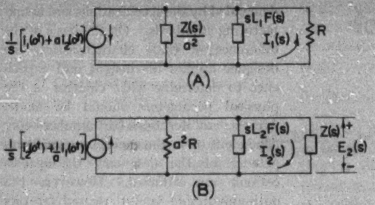


Fig. 5. Views of equivalent circuit

A—From primary (N_1) side
B—From secondary (N_2) side
a = turns ratio = N_2/N_1

By defining self- and mutual inductances as follows:

$$L_1=\frac{\mu N_1^2A}{r};\ M=\frac{\mu N_1N_2A}{r};\ L_2=\frac{\mu N_2^2A}{r}$$

and by the use of condition 35, equations 34 can be expressed as

$$RI_1(s)+\mu N_1As\bar{H}_{\text{avg}}(s)=L_1i_1(0^+)+Mi_2(0^+)$$

$$Z(s)I_2(s)+\mu N_2As\bar{H}_{\text{avg}}(s)=L_2i_2(0^+)+Mi_1(0^+)\quad(37)$$

It is necessary now to express $\bar{H}_{\text{avg}}(s)$ in terms of $I_1(s)$ and $I_2(s)$. This can be accomplished using the analysis following equation 9 with

$$\frac{Ni(0^+)}{r}$$

and

$$\frac{NI(s)}{r}$$

being replaced by

$$\frac{N_1i_1(0^+)+N_2i_2(0^+)}{r}$$

and

$$\frac{N_1I_1(s)+N_2I_2(s)}{r}$$

respectively. Referring to equation 23, the expression for $\bar{H}_{\text{avg}}(s)$ for the 2-mesh problem becomes

$$\bar{H}_{\text{avg}}(s)=\frac{N_1i_1(0^+)+N_2i_2(0^+)}{r}+\left[\frac{N_1I_1(s)+N_2I_2(s)}{r}-\frac{N_1i_1(0^+)+N_2i_2(0^+)}{rs}\right]F(s)\quad(38)$$

Equations 37 can then be written as

$$[R+sL_1F(s)]I_1(s)+sMF(s)I_2(s)=L_1F(s)i_1(0^+)+MF(s)i_2(0^+)$$

$$sMF(s)I_1(s)+[Z(s)+sI_2F(s)]I_2(s)=MF(s)i_1(0^+)+L_2F(s)i_2(0^+)\quad(39)$$

Solution of these equations gives the following expressions for the transforms of $i_1(t)$, $i_2(t)$, and $e_2(t)$:

$$I_1(s)=\frac{Z(s)F(s)[L_1i_1(0^+)+Mi_2(0^+)]}{Z(s)[R+sL_1F(s)]+RsL_2F(s)}\quad(40)$$

$$I_2(s)=\frac{RF(s)[Mi_1(0^+)+L_2i_2(0^+)]}{Z(s)[R+sL_1F(s)]+RsL_2F(s)}\quad(41)$$

$$E_2(s)=\frac{Z(s)RF(s)[Mi_1(0^+)+L_2i_2(0^+)]}{Z(s)[R+sL_1F(s)]+RsL_2F(s)}\quad(42)$$

It is clear that the pole locations in all three transforms depend not only on the external circuit impedances R and $Z(s)$, but also on the characteristics of the laminations whose properties enter through the function $F(s)$. From the equations, equivalent circuits may be determined with reference to either the primary (N_1) or secondary (N_2) side of the 2-mesh circuit. Figs. 5(A) and 5(B) illustrate the form of the equivalent circuit representation. The effect of the coils wound on the magnetic material is taken into account in circuit calculations by the elements $sL_1F(s)$ and $sL_2F(s)$. In the absence of eddy currents these elements would have the value sL_1 and sL_2.

Conclusions

As demonstrated in the specific example discussed, iterated Laplace transformations have considerable value in the solution of combined circuit-field problems In problems of this kind where partia differential equations arise whose boundary conditions contain unknowns appearing as variables in associated ordinary differential equations, the iterated transformations permit solutions to the complete problem to be obtained readily. Furthermore, in the particular problem analyzed here, the iterated transform method allowed the determination of a simple equivalent circuit representation for the physical system in a straightforward fashion.

References

1. MODERN OPERATIONAL MATHEMATICS IN ENGINEERING (book), **Ruel V. Churchill.** McGraw-Hill Book Company, Inc., New York, N. Y., 1944, pp. 210–11.

2. OPERATORENRECHNUNG NEBST ANWENDUNG IN PHYSIK UND TECHNIK (book), **Karl Willy Wagner.** Edwards Brothers, Ann Arbor, Mich., 1944, pp. 233–37.

3. UBER DIE WIRBELSTROMVERZOGERUNG MAGNETISCHER SCHALTVORGANGE, **W. Wolman, H. Kaden.** *Zeitschrift fur Technische Physik*, Leipzig, Germany, vol. 13, no. 7, 1932, pp. 330–35.

4. APPLICATION OF THIN PERMALLOY TAPE IN WIDE BAND TELEPHONE AND PULSE TRANSFORMERS, **A. G. Ganz.** *AIEE Transactions*, vol. 65, Apr. 1946, pp. 177–83.

5. PENETRATION OF AN ELECTROMAGNETIC WAVE INTO A FERROMAGNETIC MATERIAL, **A. Papoulis.** *Journal of Applied Physics*, New York, N. Y., vol. 25, no. 2, Feb. 1954, pp. 169–76.

6. EDDY CURRENT PHENOMENA IN FERROMAGNETIC MATERIALS, **H. M. McConnell.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 73, July 1954, pp. 226–35.

7. FLUX REVERSAL IN MAGNETIC AMPLIFIER CORES, **F. J. Friedlander.** *Ibid.*, vol. 75, July 1956, pp. 268–77.

8. THEORY AND DESIGN OF HIGH POWER PULSE TRANSFORMERS, **W. S. Melville.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 93, 1946, pt. III-A, pp. 1063–80.

9. THE MEASUREMENT AND CALCULATION OF PULSE MAGNETIZATION CHARACTERISTICS OF NICKEL IRONS FROM 0.1 TO 5 MICROSECONDS. *Ibid.*, vol. 97, 1950, pt. II, pp. 165–98.

10. MAGNETIC MATERIALS FOR DIGITAL COMPUTER COMPONENTS. II—MAGNETIC CHARACTERISTICS OF ULTRA-THIN MOLYBDENUM-PERMALLOY CORES, **Norman Menyuk.** *Journal of Applied Physics*, vol. 26, no. 6, June 1955, pp. 692–97.

11. TERMINAL PROPERTIES OF MAGNETIC CORES, **T. C. Chen, A. Papoulis.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 46, no. 5, pt. I, May 1958, pp. 839–49.

12. PRINCIPLES AND TECHNIQUES OF APPLIED MATHEMATICS (book), **Bernard Friedman.** John Wiley & Sons, Inc., New York, N. Y., 1956, chap. 3.

System Synthesis with the Aid of Digital Computers

J. B. DENNIS
NONMEMBER AIEE

R. F. NEASE
NONMEMBER AIEE

R. M. SAUNDERS
MEMBER AIEE

Synopsis: The conventional procedure for designing systems with many physical parameters is discussed and the translation of this procedure into digital computer operations is considered. Although it is possible to code all of the routine design calculations for a computer, it is impossible to program the insight of the design engineer. It is shown that the determination of the physical parameters for the optimum design of a system is generally equivalent to the minimization of a function subject to constraints. The application of the method of steepest descent to minimization problems is discussed. A method for extending its use to minimization with constraints is presented.

FOR YEARS, the designer responsible for synthesizing aircraft electrical systems developed his techniques from the methods of electric machine design. By using these methods, he has developed system designs chiefly by repetitive calculation and considerable insight. The existence of systems of high caliber show how successful many designers have been. With the current trend toward larger and more complex systems, however, the designer is being requested to acquire considerably more insight than was previously required. Hence it is essential to search for more efficient system design procedures than are currently in use.

Analytic techniques are available for certain classes of design problems. For instance, methods for synthesizing linear static networks are well developed, permitting one to find the circuit parameters to meet a given set of specifications.[1] In the case of complex system designs, analytic techniques are generally not applicable and thus numerical techniques must be used. The modern digital computer offers a practical means of solving such system design problems if an efficient method for their solution is available. Considerable effort is being expended on the application of a digital computer to the design of transformers.[2] This paper describes a numerical technique which provides an approach for obtaining the optimum design for a class of nonlinear systems once the system configuration is known.

The research reported here was undertaken to study the application of digital computers specifically to the design of an aircraft alternator and its associated excitation-regulation system. However, the techniques should be applicable to many design problems in other fields as well as in electrical engineering.

Conventional Design Procedure

The design of electrical systems to meet given specifications and to have minimum weight, cost, or volume is generally carried out by a trial and error procedure. This conventional design procedure is shown in in block diagram form in Fig. 1. At the beginning of a new design, the designer makes a choice of physical parameters either by assumption or from a known design. He then compares their resulting performance characteristics with those demanded by the specifications and manufacturing standards. On the basis of the difference of these characteristics, the designer applies his insight and experience to determine what changes in the physical parameters should be made. After several iterations (the number being dependent upon the experience and competence of the designer), the specifications are satisfied. However, the optimum design is not assured by this process. Numerous other designs must be made to assure that a better system design is not overlooked. At the end of this process, the most desirable system which meets all specifications, is available.

Although the calculations of the performance parameters from assumed physical parameters may be speeded up considerably by the use of a digital computer the entire conventional design procedure is not speeded up appreciably. This is true because the decisions based upon insight would be impossible to state mathematically. However, there are procedures which may be carried out entirely on a digital computer. Such a procedure is the topic of this paper. It is to be emphasized here that the authors do not propose this as a substitute for the conventional design procedure if the calculations are to be carried out manually. This procedure should be useful only if a high-speed digital computer is available for design purposes.

Characteristics of a Design Problem

In order that design formulas applicable to a certain system be found, the "topology" of the system must be known. The meaning of "topology" is illustrated by an example in which an a-c generator comprises the system. The machine may be

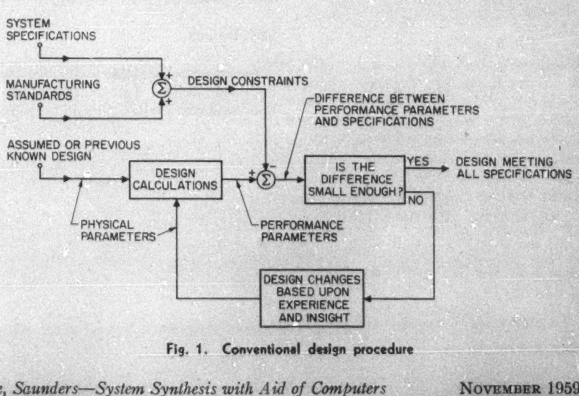

SYSTEM SPECIFICATIONS
MANUFACTURING STANDARDS
+
+
Σ
DESIGN CONSTRAINTS
DIFFERENCE BETWEEN PERFORMANCE PARAMETERS AND SPECIFICATIONS
ASSUMED OR PREVIOUS KNOWN DESIGN
DESIGN CALCULATIONS
+
−
Σ
IS THE DIFFERENCE SMALL ENOUGH?
YES
NO
DESIGN MEETING ALL SPECIFICATIONS
PHYSICAL PARAMETERS
PERFORMANCE PARAMETERS
DESIGN CHANGES BASED UPON EXPERIENCE AND INSIGHT

Fig. 1. Conventional design procedure

Paper 59-196, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 30, 1955; made available for printing December 3, 1958.

J. B. DENNIS is with the Massachusetts Institute of Technology (M. I. T.), Cambridge, Mass.; R. F. NEASE is with North American Aviation, Inc., Los Angeles, Calif.; and R. M. SAUNDERS was Visiting Associate Professor of Electrical Engineering at M. I. T. on leave from the University of California at Berkeley, Calif., during the period of time covered by this paper.

This work was conducted by the Energy Conversion Group of the Servomechanisms Laboratory at M. I. T. and was supported, in part, by Wright Air Development Center under Contract AF 33(616)-2609.

Vol.
Pt. 1
195

an asynchronous generator or a synchronous generator. The type of rotor structure must also be specified. The fixing of such features as these are referred to in this paper as the specification of the topology of the system.

After fixing the topology of the system, design formulas may be chosen that characterize the system operation and thus provide the design relations for a particular application. The design problem is then as follows. First, the specifications on system performance must be stated in terms of performance parameters given by the design formulas. If an alternator comprises the system, one constraint on a performance parameter could be that the transient reactance be less than a specified value. Another constraint might be that the total losses of the alternator be less than a fixed amount. After determining which performance parameters are to be constrained in the design, the physical parameters which are important in meeting the specifications are then available by inspection of the design formulas.

In addition, the criterion used to define optimum must be known to complete the statement of the design problem. In the case of the design of an aircraft alternator, this criterion could be the minimization of the machine weight. In other design problems, the criterion might be to achieve the minimum cost or volume for the system. Also, the minimization of a combination of such factors might be used as the criterion.

Mathematical Representation of the Design Problem

In the previous discussion, the design problem has been treated as one of meeting given specifications and optimizing in terms of a given criterion. The topology of the system to be designed is assumed fixed. The mathematical statement of the design problem is as follows. Let the quantities $x_1, x_2, \ldots, x_m$ be the physical parameters of the system which the designer may vary. The fact that n specifications must be met may be stated as

$$\begin{aligned} &f_1(x_1, x_2, \ldots, x_m) \leq y_1 \\ &f_2(x_1, x_2, \ldots, x_m) \leq y_2 \\ &\quad \cdot \qquad \cdot \qquad \cdot \\ &\quad \cdot \qquad \cdot \qquad \cdot \\ &\quad \cdot \qquad \cdot \qquad \cdot \\ &f_n(x_1, x_2, \ldots, x_m) \leq y_n \end{aligned} \tag{1}$$

where the functions $f_1, f_2, \ldots, f_n$ are the design formulas giving performance parameters in terms of the physical parameters. The constants $y_1, y_2, \ldots, y_n$ are fixed by the specifications. The criterion for optimization may be written as the minimization of a function of the physical parameters

$$G = G(x_1, x_2, \ldots, x_m) \tag{2}$$

Thus the design problem may be stated as the minimization of a function (equation 2) in the presence of constraints (equation 1). If the functional relations of equation 1 were available in analytical form, the use of Lagrange multipliers[3] would provide a straightforward method of minimization. However, design formulas for systems usually will involve empirical functions which are not amenable to analytical treatment. For this reason and due to the usual complexity of design formulas, the Lagrange method is not too useful for design in general. The following sections present a numerical method for minimizing a function in the presence of constraints, where the constraint relations need not be given as analytical expressions in terms of the physical parameters of the system. This method is readily adaptable for use with a digital computer.

The Method of Steepest Descent

The problem of minimizing a function of several variables in the absence of constraints will be considered first. This technique is known as the method of steepest descent[4].

For purposes of illustration, consider a function, $G(x_1, x_2)$, of the two independent variables, x_1 and x_2. It is desired to minimize G with respect to x_1 and x_2; that is, find the point in the x_1x_2 plane for which $G(x_1,x_2)$ takes on the smallest value. Let P_m in Fig. 2 be the point at which $G(x_1,x_2)$ is a minimum. If one moves in any direction from P_m, the value of $G(x_1,x_2)$ must increase. Therefore, near P_m, the lines along which $G(x_1, x_2)$ is constant must be closed curves encircling P_m. Let the point P_0 be specified by initial values of the independent variables,

$$\begin{aligned} x_1 &= x_1^0 \\ x_2 &= x_2^0 \end{aligned} \tag{3}$$

The choice of these initial values will be discussed later. A new set of values for x_1 and x_2 is found by moving from P_0 in the direction that gives the largest decrease of $G(x_1, x_2)$ per unit distance travelled in the x_1x_2 plane. This is the direction of the vector gradient of $-G(x_1, x_2)$ and is a direction perpendicular to the line of constant $G(x_1,x_2)$ at P_0. Thus the values of the variables at the new point P_1 are given by:

$$\begin{aligned} x_1^1 &= x_1^0 - \lambda \left.\frac{\partial G}{\partial x_1}\right|_{x_1 = x_1^0,\, x_2 = x_2^0} \\ x_2^1 &= x_2^0 - \lambda \left.\frac{\partial G}{\partial x_2}\right|_{x_1 = x_1^0,\, x_2 = x_2^0} \end{aligned} \tag{4}$$

where λ is a positive constant. Using vector notation

$$\vec{x}^1 = \vec{x}^0 - \lambda \nabla G(x_1, x_2)\Big|_{\vec{x} = \vec{x}^0} \tag{5}$$

where

$$\nabla G = \text{gradient } G \equiv \left[\frac{\partial G}{\partial x_1}, \frac{\partial G}{\partial x_2}\right]$$

The elements, $\partial G/\partial x_1$ and $\partial G/\partial x_2$, of the matrix are the components of the vector ∇G in the x_1 and x_2 directions. It is evident that one step in the direction of the gradient will not, in general, take one to the minimum at P_m. Hence, this procedure should be repeated until some criterion is satisfied. In Fig. 2, for example, it takes six iterations to converge on P_m within the allowable error. A suitable criterion is that the change in $G(x_1,x_2)$ observed on the last step be less than some chosen value. The value of the constant λ must be controlled carefully to avoid so large a value that the solution does not converge to the point P_m. Also λ should not be so small that an excess number of iterations are required. One method of controlling λ is by observing the angle between successive evaluations of the gradient vector. If the angle is smaller than a given value, λ may be doubled; if the angle is too large, λ is halved. In addition λ is never allowed to be so large that $G(x_1,x_2)$ is increased.

This method may be extended to any number of independent variables with only the loss of the simple geometric interpretation given above.

Minimization with Constraints

The problem of minimizing a function of m variables subject to n equality constraints is considered first. That is, the

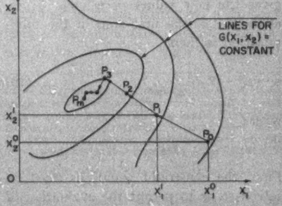


Fig. 2. Method of steepest descent

equality signs hold in equation 1. In general, only n of the variables, say $x_1, x_2, \ldots, x_n$, are constrained by equation 1. Therefore, k of the variables where $k=m-n$, may be changed so as to achieve the minimization of G. These k variables will be called $w_1, w_2, \ldots, w_k$. The problem then becomes one of minimizing $G=G(x_1, x_2, \ldots, x_n, w_1, w_2, \ldots, w_k)$ subject to the constraints

$$
\begin{aligned}
&f_1(x_1, x_2, \ldots, x_n, w_1, w_2, \ldots, w_k)=y_1\\
&\quad\vdots\\
&f_n(x_1, x_2, \ldots, x_n, w_1, w_2, \ldots, w_k)=y_n
\end{aligned}
\qquad (6)
$$

Since the variables $x_1, x_2, \ldots, x_n$ are now dependent, G is an implicit function of $w_1, w_2, \ldots, w_k$; that is,

$$G=G(w_1, w_2, \ldots, w_k) \qquad (7)$$

If this function could be obtained explicitly by elimination of $x_1, x_2, \ldots, x_n$ by the use of equation 6, the problem could be solved by the minimization procedure for the case of no constraints. However, an explicit expression for equation 7 would in general be impossible to obtain. On the other hand, it is not really necessary to have an explicit form of equation 7 to be able to apply the method of steepest descent to the problem. One needs only to evaluate the derivatives

$$
\begin{aligned}
&\left.\frac{\partial G}{\partial w_1}\right|_f \simeq \left.\frac{\Delta G}{\Delta w_1}\right|_{\substack{w_2, w_3, \ldots, w_k \text{ constant}\\ \Delta f_1=\Delta f_2=\ldots=\Delta f_n=0}}\\
&\quad\vdots\\
&\left.\frac{\partial G}{\partial w_k}\right|_f \simeq \left.\frac{\Delta G}{\Delta w_k}\right|_{\substack{w_1, w_2, \ldots, w_{k-1} \text{ constant}\\ \Delta f_1=\Delta f_2=\ldots=\Delta f_n=0}}
\end{aligned}
\qquad (8)
$$

in order to evaluate the gradient of G

$$\nabla_f G \equiv \left[\left.\frac{\partial G}{\partial w_1}\right|_f, \left.\frac{\partial G}{\partial w_2}\right|_f, \ldots, \left.\frac{\partial G}{\partial w_k}\right|_f\right] \qquad (9)$$

The partial derivatives of equation 8 are written in terms of differences to indicate the numerical technique used for their evaluation.

The equivalent of equation 5 which gives the new values of $w_1, w_2, \ldots, w_k$ at the end of the first step in the minimization is then

$$\vec{w}^1=\vec{w}^0-\lambda\nabla_f G\Big|_{\vec{w}=\vec{w}^0} \qquad (10)$$

where

$$\vec{w}=[w_1, w_2, \ldots, w_k]$$

First, the constraint equations must be solved so as to permit the evaluation of $\nabla_f G$. Let $\Delta x_{1,1}, \Delta x_{2,1}, \ldots, \Delta x_{n,1}$ be the changes in $x_1, x_2, \ldots, x_n$ respectively, necessary to keep the constraints satisfied when w_1 has been changed by a small amount Δw_1 while the other w's remain constant. These Δx's are obtained by solving the following set of linear equations in terms of Δw_1.

$$
\begin{aligned}
&-\frac{\partial f_1}{\partial w_1}\Delta w_1 \simeq \frac{\partial f_1}{\partial x_1}\Delta x_{1,1}+\frac{\partial f_1}{\partial x_2}\Delta x_{2,1}+\ldots+\frac{\partial f_1}{\partial x_n}\Delta x_{n,1}\\
&\quad\vdots\\
&-\frac{\partial f_n}{\partial w_1}\Delta w_1 \simeq \frac{\partial f_n}{\partial x_1}\Delta x_{1,1}+\frac{\partial f_n}{\partial x_2}\Delta x_{2,1}+\ldots+\frac{\partial f_n}{\partial x_n}\Delta_{n,1}
\end{aligned}
\qquad (11)
$$

Equation 11 is obtained from equation 6 by setting $\Delta f_1=\Delta f_2=\ldots\Delta f_n=0$. The partial derivatives of equation 11 are evaluated at the point $\vec{w}=\vec{w}^0$ and are constants. Therefore, the solution is of the form

$$
\begin{aligned}
&\Delta x_{1,1}\simeq\alpha_{1,1}\Delta w_1\\
&\quad\vdots\\
&\Delta x_{n,1}\simeq\alpha_{n,1}\Delta w_1
\end{aligned}
\qquad (12)
$$

where $\alpha_{1,1}, \ldots, \alpha_{n,1}$ are constants. Since by definition

$$\left.\frac{\partial G}{\partial w_1}\right|_f \equiv \lim_{\Delta w_1\to 0}\frac{\dfrac{\partial G}{\partial x_1}\Delta x_{1,1}+\ldots+\dfrac{\partial G}{\partial x_n}\Delta x_{n,1}+\dfrac{\partial G}{\partial w_1}\Delta w_1}{\Delta w_1} \qquad (13)$$

then from equation 12,

$$\left.\frac{\partial G}{\partial w_1}\right|_f \simeq \frac{\partial G}{\partial x_1}\alpha_{1,1}+\ldots+\frac{\partial G}{\partial x_n}\alpha_{n,1}+\frac{\partial G}{\partial w_1} \qquad (14)$$

In the same manner,

$$\left.\frac{\partial G}{\partial w_2}\right|_f, \ldots, \left.\frac{\partial G}{\partial w_k}\right|_f$$

may be evaluated and hence $\vec{w}^1$ is given by equation 10. However, the new values of the dependent variables $x_1{}^1, x_2{}^1, \ldots, x_n{}^1$ must be found before another step toward the minimum of G is made. Since the change in x_1 in moving from P_0 to P_1 is

$$x_1{}^1-x_1{}^0=\alpha_{1,1}(w_1{}^1-w_1{}^0)+\alpha_{1,2}(w_2{}^1-w_2{}^0)+\ldots+\alpha_{1,k}(w_k{}^1-w_k{}^0), \qquad (15)$$

$$x_1{}^1=x_1{}^0-\lambda\left[\alpha_{1,1}\left.\frac{\partial G}{\partial w_1}\right|_f+\alpha_{1,2}\left.\frac{\partial G}{\partial w_2}\right|_f+\ldots+\alpha_{1,k}\left.\frac{\partial G}{\partial w_k}\right|_f\right] \qquad (16)$$

by combining equations 10 and 15, thereby giving the value of x_1 at the new point P_1. In the same way, the values of $x_2, \ldots, x_n$ at P_1 are found. Then, the same procedure is initiated again to evaluate the next step toward the minimum, and so on until the point P_m is reached The value of λ can be controlled in a manner identical to the method used for the unconstrained case described in the previous section.

This method for minimizing a function in the presence of equality constraints is readily applicable to digital computation. The partial derivatives are approximated by the use of difference methods and the solution of the linear simultaneous equations of equation 11 may be effected by standard matrix inversion techniques.

An extension of this method, which can be used in solving a design problem when the constraints are inequalities, is as follows. (Since this paper was written J. B. Dennis has developed a general procedure for minimizing a function subject to nonlinear inequality constraints. This method will always converge to the nearest relative minimum. See reference 5.) A set of initial values of the physical parameters is chosen such that equation 1 is satisfied. Since the inequalities hold for all of the constraints, then each of the physical parameters may be changed at least a small amount without violating any of the constraint equations. Thus, in the beginning, the problem may be treated as minimization without constraints. If a minimum is obtained without ever violating any of the constraint equations, the constraints are not significant or the phrasing of the problem is inconsistent. However, if any of the constraints are significant, one of them will eventually be violated as steps toward the minimum are taken. At the point where a constraint is first violated its equal sign is assumed to hold and the problem becomes one of minimizing with one equality constraint which is treated by the procedure previously described. Thus, the number k of independent variables w is decreased by one, and the number n of dependent variables is one. As more steps are made, more of the constraint equations will be violated and their effects brought into the computation by the addition of dependent variables until the minimum is reached.

Difficulties with the Steepest Descent Method

Since the method of steepest descent described involves taking derivatives, it is basically applicable to problems for which the functions involved are continuous and have continuous derivatives. In many design problems there are both continuous variables to which any value may be assigned, and discrete variables which may only take on certain fixed values. The discrete variables can be handled with the minimization procedure in one of two ways depending on the number of values each discrete variable is allowed to assume. If there are only a small number of values that a particular discrete variable may have, then the design problem is solved by applying the minimization process once for each value of the discrete variable and noting which one gives the smallest minimum of G. If there are many possible values for the discrete variable this method becomes impractical because of computer time consumption. For these cases, one can replace the discrete variable with a continuous variable such that the necessary derivatives would be continuous and sufficiently smooth. Then the minimization could be performed resulting in a value of the continuous variable lying between two allowed values of the original discrete variable. These two values could then be tried separately and the better one chosen for the final design. These ideas can be extended to design problems involving several discrete variables.

A more serious difficulty with the steepest descent method is that the minimization problem may have several minima. The minimization procedure will find one of these minima, the exact one depending on the choice of initial values of the variables. A check can be made as to whether the true minimum had been obtained by trying several widely differing initial choices of parameter values and verifying that the same answer is obtained. Another way is to insure that the initial selection of parameter values is sufficiently close to the optimum so that there would be no difficulty. Such a set of physical parameters may be known from a previous design, or may be obtained by preliminary calculations.

Conclusions

This paper provides a numerical method (adaptable to a digital computer) for minimizing a function in the presence of constraints.

The method should prove useful in design optimization is the design problem can be expressed subject to the assumptions which follows.

1. The topology of the system is fixed.
2. Design formulas (relations between performance parameters and the physical parameters) are available.
3. The process of optimization is equivalent to the minimization of some function of the physical parameters.
4. The specifications may be expressed in terms of n constraint inequalities involving functions of the physical parameters.
5. An initial design is available which satisfied the constraint equations.
6. Only one minimum exists for the function being minimized or a sufficiently good initial design is known.

References

1. Automatic Feedback Control System Synthesis (book), **J. G. Truxal.** McGraw-Hill Book Company, Inc., New York, N. Y., 1955, chap. 5.

2. How Digital Computers Aid Transformer Designers, **S. B. Williams, P. A. Abetti, E. F. Magnusson.** *General Electric Review*, Schenectady, N. Y., vol. 58, May 1955, pp. 24–25.

3. Higher Mathematics for Engineers (book), **I. E. and E. S. Sokolnikoff.** McGraw-Hill Book Company, Inc., New York, N. Y., 1941, second edition, pp. 163–66.

4. Principles of Numerical Analysis (book), **A. S. Householder.** McGraw-Hill Book Company, Inc., New York, N. Y., 1953, pp. 132–35.

5. Mathematical Programming and Electrical Networks, **J. B. Dennis.** John Wiley & Sons, Inc., New York, N. Y., 1959, Chap. 7.

The Numerical-Graphical Method in the Design of Multiterminal Switching Circuits

A. H. SCHEINMAN
ASSOCIATE MEMBER AIEE

UP TO the present time there has been no satisfactory direct method for designing multiterminal switching circuits. A common approach is to consider each output individually and derive a 2-terminal network for it. An attempt is then made to change the form of the individual 2-terminal networks, if possible, so that combinations between paths may be facilitated. This may be done, for example, by algebraic manipulation of the switching functions describing the 2-terminal networks. Extreme care must be taken in any case so that the individual outputs remain disjunctive.

The type of multiterminal circuit generally encountered is illustrated in Fig. 1. This circuit has a multiplicity of input leads and two or more output leads. The input leads control relays which serve to provide the necessary contacts to form the paths from some common point, usually ground, to control the output leads in accordance with the circuit specifications.

Two methods are presented in this paper for designing multiterminal circuits. Both methods are based upon the numerical-graphical method[1] which will now be summarized.

In the switching algebra that appears in this paper, the following conventions are used:

0 = open circuit.
1 = closed circuit.
Addition (+) = parallel connection = or.
Multiplication = series connection = and.
An unprimed variable denotes a make contact.
A primed variable denotes a break contact.

Procedure Summarized[1]

The desired switching function can be represented by a table of combinations, each row of which represents a product term of the disjunctive canonical expansion of the function. Each of the input

---

Paper 59-1028, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for publication without prior presentation. Manuscript submitted October 28, 1958; made available for printing May 13, 1959.

A. H. Scheinman is with Bell Telephone Laboratories, Inc., New York, N. Y.

This paper represents material submitted to the Polytechnic Institute of Brooklyn, Brooklyn, N. Y., in partial fulfillment of the requirements for the degree of Master of Electrical Engineering.

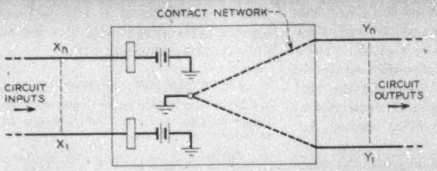

CONTACT NETWORK
X_n
X_1
CIRCUIT INPUTS
CIRCUIT OUTPUTS
Y_n
Y_1

Fig. 1. General multiterminal network

variables has a binary weight associated with it depending on its position in the table. Each row can then be represented by a decimal equivalent which enables the original switching function to be conveniently specified by a series of decimal numbers.

A transfer contact of the most heavily weighted variable is then introduced. Two new functions are formed, one in series with the break contact and one in series with the make contact. The function in series with the break contact consists of all of the decimal numbers of the original function which are less than the weight of the variable introduced. The function in series with the make contact consists of the decimal numbers which are equal to or greater than the weight of the variable introduced, from each of which the weight of the variable is subtracted. The new functions are compared to see if one of the following rules may be applied:

Rule 1. Equal functions of the same variables may be combined.

Rule 2. If one of the two functions emanating from a transfer contact is a subset of the other, the element of the transfer contact in series with the subset may be short circuited. The decimal numbers in series with the other element of the transfer contact which correspond to those of the subset become "don't-care" terms.

Rule 3. If in the development of the circuit identical contact configurations appear in series with the make and break elements of the transfer contact being introduced, they may be shifted to the left of the transfer contact and combined.

Multiterminal Circuits

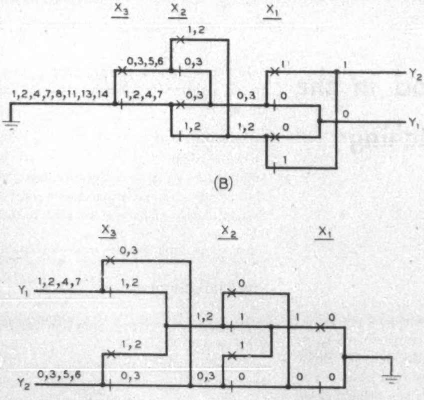
Two approaches to the design of multiterminal circuits will now be presented. The three rules summarized before, of the numerical-graphical method apply in the same manner as for 2-terminal circuit design.

Method 1

In this method a single decimal number identifies a product term of the canonical expansion and in addition the output lead

8 X_3	4 X_2	2 X_1	1 Y_2	0 Y_1	DECIMAL EQUIV.
0	0	0	1	0	1
0	1	1	1	0	7
1	0	1	1	0	11
1	1	0	1	0	13
0	0	1	0	1	2
0	1	0	0	1	4
1	0	0	0	1	8
1	1	1	0	1	14

(A)


X_3
X_2
X_1
1,2
0,3,5,6
0,3
1,2,4,7,8,11,13,14
1,2,4,7
0,3
0,3
0
1
1
Y_2
0
Y_1
1,2
1,2
0
1
X_3
X_2
X_1
0,3
Y_1
1,2,4,7
1,2
0
1,2
1
1
0
1,2
1
Y_2
0,3,5,6
0,3
0,3
0
0
0

(B)

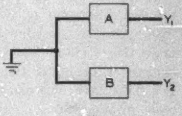
(C)

Fig. 2. Two methods of multiterminal-circuit design

A—Table of combinations
B—Circuit development for method 1
C—Circuit development for method 2

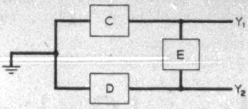
X_3	X_2	X_1	Y_2	Y_1
0	0	0	0	1
0	0	1	1	0
0	1	0	1	1
0	1	1	1	1
1	0	0	0	1
1	0	1	1	0

(A)



(B)

(C)



(D)

	8 X_3	4 X_2	2 X_1	1 Y_2	0 Y_1	DECIMAL EQUIV.
	0	0	0	0	1	0
	0	0	1	1	0	3
E	0	1	0	0	1	4
	0	1	1	1	0	7
	1	0	0	0	1	8
	1	0	1	1	0	11

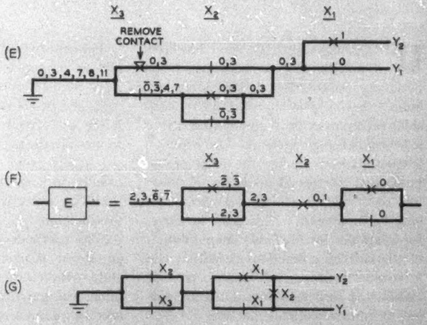
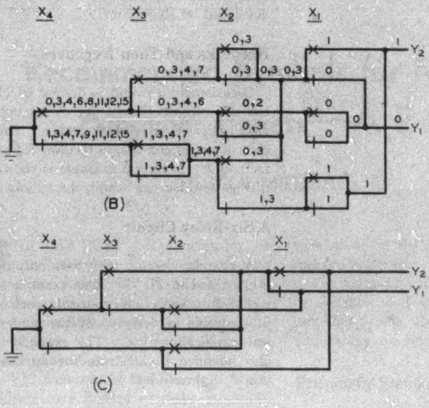

(E)
X_3
X_2
X_1
REMOVE CONTACT
0,3,4,7,8,11
0,3
0,3
0,3
1
Y_2
0
Y_1
$\bar{0},\bar{3},4,7$
0,3
0,3
$\bar{0},\bar{3}$
(F)
X_3
X_2
X_1
E
=
$2,3,\bar{6},\bar{7}$
$\bar{2},\bar{3}$
2,3
2,3
0,1
0
0
(G)
X_2
X_3
X_1
X_1
X_2
Y_2
Y_1

Fig. 3 (right). Extension of method 1

Vol.
Pt. I
195

16 X_4	8 X_3	4 X_2	2 X_1	1 Y_2	0 Y_1	DECIMAL EQUIV.
0	0	1	0	0	1	4
0	1	1	0	0	1	12
1	0	0	0	0	1	16
1	0	1	0	0	1	20
1	0	1	1	0	1	22
1	1	0	0	0	1	24
1	1	1	0	0	1	28
0	0	0	0	1	0	1
0	0	0	1	1	0	3
0	0	1	1	1	0	7
0	1	0	0	1	0	9
0	1	0	1	1	0	11
0	1	1	1	1	0	15
1	0	0	1	1	0	19
1	1	0	1	1	0	27
1	1	1	1	1	0	31

(A)



64 X_6	32 X_5	16 X_4	8 X_3	4 X_2	2 X_1	1 Y_2	0 Y_1	DECIMAL EQUIV.
1	1	1	1	1	1	0	1	126
1	1	1	0	0	1	0	1	114
1	1	0	1	0	1	0	1	106
1	0	1	1	0	1	0	1	90
0	0	1	1	1	1	0	1	30
0	0	1	0	0	1	0	1	18
0	0	0	1	0	1	0	1	10
1	1	1	1	0	0	1	0	121
1	0	1	1	1	0	1	0	93
1	0	0	1	1	0	1	0	77
0	0	1	1	0	0	1	0	25

(A)

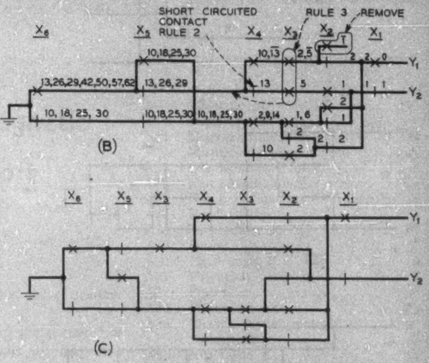


Fig. 4 (left). Development of network and its negative

A—Table of combinations
B—Circuit development
C—Completed circuit

Fig. 5 (above). Method 1 with six relays. See Fig. 4 caption for identifications

to which a closure must be made if that combination of input variables occurs. A simple example will make this clear.

Assume a circuit with 3 inputs, X_1, X_2, and X_3, and 2 outputs, Y_1 and Y_2, such that Y_1 and Y_2 are grounded according to the following equations:

$$f(Y_1)=X_3'X_2'X_1+X_3'X_2X_1'+X_3X_2'X_1'+X_3X_2X_1$$

$$f(Y_2)=X_3'X_2'X_1'+X_3'X_2X_1+X_3X_2'X_1+X_3X_2X_1'$$

A composite table of combinations is drawn as shown in Fig. 2(A), where a 1 in an output column indicates that the output is grounded for the combination of input variables in the corresponding row. For example, the combination X_3' X_2' X_1' causes output Y_2 to be grounded, and this is indicated by the first row of the table of combinations, 00010.

The entire row is then assumed to be one binary number and is replaced by its decimal equivalent.

There is one restriction for the direct use of method *1*. For any particular combination of input variables only one output may be grounded. Taking advantage of this restriction the first output is arbitrarily given a weight of zero instead of one as it would normally be in a true binary number. This is done to make the 2-terminal circuit a special case of method *1* where the single output is given zero weight. It is of course just as valid to consider the number as a true binary and give the first output a weight of one. The former scheme will be adopted here however.

The decimal equivalents are obtained as shown in Fig. 2(A). Fig. 2(B) shows the development of the circuit, which is carried out exactly as for the 2-terminal case. The difference becomes apparent after transfers of the last variable are introduced. There will then be as many leads as there are outputs and the leads will be numbered the numbers corresponding to the weights of the outputs. In Fig. 2(B) the two output leads are numbered 0 and 1. The table of Fig. 2(A) shows that 0 is the weight of Y_1 and 1 is the weight of Y_2, thus identifying the outputs. The circuit has now been completed and we note that five transfer contacts are required.

As noted previously, if more than one output is to be grounded for any particular combination of input variables, then method *1* cannot be used directly. However, by using a bridging network across the outputs in question consisting of the common combinations of input variables, the problem can be converted to one in which method *1* can be used.

Consider as an example a circuit as shown by the table of combinations of Fig. 3(A).

$$f(Y_1)=X_3'X_2'X_1'+X_3'X_2X_1'+X_3'X_2X_1+X_3X_2'X_1'$$

$$f(Y_2)=X_3'X_2'X_1+X_3'X_2X_1'+X_3'X_2X_1+X_3X_2'X_1$$

This is shown symbolically in Fig. 3(B) where $f(Y_1)$ is shown as network *A* and $f(Y_2)$ as network *B*. Since some of the combinations of networks *A* and *B*

4	2	1			DECIMAL EQUIV. OF
X_3	X_2	X_1	Y_2	Y_1	INPUT VARIABLES
1	1	1	0	1	7
0	1	1	1	0	3
1	1	0	0	1	6
0	1	0	1	0	2
1	0	1	1	1	5
0	0	1	1	1	1

(A)

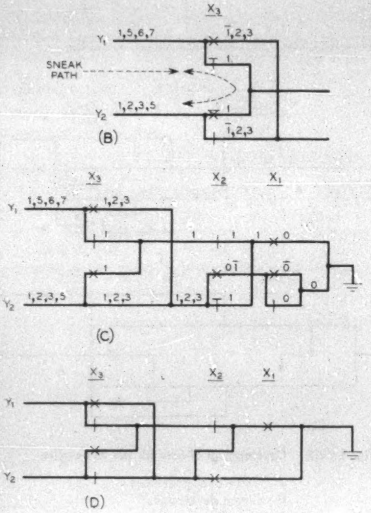


Fig. 6. Circuit development by method 2 showing sneak path

A—Table of combinations
B—Sneak path
C—Circuit development
D—Completed circuit

are common to both, method *1* cannot be used directly. If, however, networks *A* and *B* are converted to networks *C* and *D* [see Fig. 3(C)] which have no common combinations and a network *E* is added which is composed of the common combinations, then method *1* can be used to synthesize a circuit composed of the *C* and *D* networks. Network *E* is a 2-terminal network which can be designed directly. Note that the common input combinations must appear in either network *C* or *D* but not both. It is often helpful in determining whether to use any particular combination in either *C* or *D* to use a Karnaugh map.[2] The choice is made in a manner consistent with the normal use of the map, that is, to obtain the smallest number of subcubes which are as large in dimensions as possible. The circuit can now be developed as shown in Fig. 3(E).

The bridging network development is shown in Fig. 3(F). Note that any combinations which are not used in either networks *C* or *D* may be included in the bridging network as don't care combinations. The completed circuit is shown in Fig. 3(G).

METHOD 2

The second method has no restrictions in its use as does method *1* and may always be applied directly. The resulting circuits, however, are not always as efficient as those obtained by the use of method *1*.

In method *2*, only the input variables are weighted and the decimal equivalents are based on the input combinations only. Thus in Fig. 2(A), to apply method *2*, the decimal equivalents are based on the input variable combinations and X_3, X_2, and X_1 have binary weights of 4, 2, and 1 respectively. Therefore:

$$f(Y_1) = \Sigma\, 1, 2, 4, 7$$

$$f(Y_2) = \Sigma\, 0, 3, 5, 6$$

The circuit is then developed from the outputs toward the common input, which is here assumed to be ground. Interconnections are made between circuits by application of rule *1*. Fig. 2(C) shows the development of the circuit and it is noted that five transfer contacts are required as in the circuit development using method *1*. The distribution of contacts differs however.

Although the total number of contacts required in this simple example are the same when using either method *1* or *2*, this will not generally be true. Use of method *1* results in circuits that are at least as economical of contacts as method *2* and often more economical. In addition the circuit economy resulting from use of method *2* is often quite sensitive to the order of variables in the table of combinations. Method *2*, as described here, is similar to a circuit design method developed by Roginskii.[3]

Networks and Their Negatives

Occasionally a circuit is desired which has two outputs, where one of the outputs is the negative of the other. This class of circuit can always be designed by method *1*. A typical example is shown in Fig. 4 which can readily be followed.

A Six-Relay Circuit

A six-relay circuit with two outputs is illustated in Fig. 5. This example is included to show how routinely such a problem can be designed by the numerical-graphical method. The example is also intended to illustrate the application of all three rules.

Extraneous Circuit Closures

Extraneous circuit closures or "sneak paths" may inadvertently be developed by application of rule *2* followed by rule *1*. If this should occur, rule *2* should not be applied.

Fig. 6 illustrates the development of a sneak path using method *2*. When using method *2*, rule *2* must be applied with caution since the sneak path may not appear immediately, but only after contacts of subsequent variables are introduced. It is usually easier to detect sneak paths when using method *1* than with method *2*. Note that sneak paths can never be developed if rule *2* is not applied.

Use of Other Codes

In this paper the decimal code was used exclusively. However, there is no reason to preclude the use of other codes since the codes are used as a convenient means of specifying circuit transmissions and the particular code used is incidental to the method. A very useful code, for example, is the octal code. This has an advantage over the decimal in the ease of translation from the binary numbers which specify the circuit transmission to the equivalent octal numbers. All the rules governing the application of the numerical-graphical method apply directly regardless of the code used.

Conclusions

The numerical-graphical method enables a designer to develop economical multiterminal circuits in a fairly simple and straightforward manner. Usually the circuit arrangements are at least as economical as those obtained by other methods.

Most circuit designers generally do not use any of the currently available design tools. This is probably true because practical problems usually involve large numbers of variables and present methods rapidly exceed the limits of practicability as the number of variables involved increases.

The numerical-graphical method may not be a panacea, but by its use a designer is at least able to approach the problem of designing large circuits in an orderly fashion.

References

1. A NUMERICAL-GRAPHICAL METHOD FOR SYNTHESIZING SWITCHING CIRCUITS, **A. H. Scheinman.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, 1957 (Jan. 1958 section), pp. 687–89.

2. THE MAP METHOD FOR SYNTHESIS OF COMBINATIONAL LOGIC CIRCUITS, **M. Karnaugh.** *Ibid.*, vol. 72, Nov. 1953, pp. 593–99.

3. GRAPHICAL METHOD FOR SYNTHESIZING CONTACT CIRCUITS, **V. N. Roginskii.** *Elektrosvyas*, Moscow, USSR, vol. 11, no. 11, 1957, pp. 82–88.

Recommended Symbols for Magnetic Amplifier Papers—A Report

AIEE COMMITTEE REPORT

FOR THE PURPOSE of increasing the uniformity of symbols in magnetic amplifier papers, Mr. Harold W. Lord, chairman of the AIEE Magnetic Amplifier Committee, asked the Theory Subcommittee to prepare a list of recommended symbols.

The members of the Theory Subcommittee were invited to make proposals, which were collected and circulated among the Theory Subcommittee members for comments. While some members were in complete agreement with the first proposal, others offered changes, and as a result, the original list was revised. Then, this revised list was circulated, but not only among the members of the Theory Subcommittee, but also among the members of the Main Magnetic Amplifier Committee for further comments. Of the letters received owing to this last circulation, only one letter expressed a wish for further changes. The proposals contained in this letter, however, were certain to cause considerable opposition among the members of the Theory Subcommittee, and hence, no further changes were made on the revised list of recommended symbols.

The following contains the revised list of recommended symbols for magnetic amplifier papers, obtained under the circumstances described in the foregoing.

Previously Standardized Symbols

The basis for the recommended symbols is formed by the symbols shown in Appendix IV of AIEE's Author's Guide page 19, also known as the American Standard Letter Symbols for Electrical Quantities.[1] These symbols are shown in Table A for scalars (designations for phasors are bold italics, bold roman for vectors).

Table A

A1.	Admittance	$\boldsymbol{Y}$
A2.	Angular frequency (angular velocity)	ω
A3.	Capacitance	C
A4.	Capacitivity (dielectric constant)	ϵ
A5.	Conductance	G
A6.	Conductivity	γ
A7.	Current	I
A8.	Damping constant	δ
A9.	Displacement flux	Ψ
A10.	Efficiency	η
A11.	Elastance	S
A12.	Electric displacement	$\mathbf{D}$
A13.	Electric intensity	$\mathbf{E}$
A14.	Electromotive force	E
A15.	Energy, work	W
A16.	Force	$\mathbf{F}$
A17.	Frequency	f
A18.	Impedance	$\boldsymbol{Z}$
A19.	Inductance	L
A20.	Magnetic intensity	$\mathbf{K}$
A21.	Magnetic flux	Φ
A22.	Magnetic induction	$\mathbf{B}$
A23.	Magnetic vector potential	$\mathbf{A}$
A24.	Magnetomotive force	$\mathcal{F}$
A25.	Mutual inductance	L_m
A26.	Number of conductors or turns	N
A27.	Permeability	μ
A28.	Permeance	$\mathcal{P}$
A29.	Period	T
A30.	Phase angle	φ
A31.	Potential	V
A32.	Power	P
A33.	Power factor	$\boldsymbol{F_p}$
A34.	Electric charge (quantity of electricity)	Q
A35.	Reactance	X
A36.	Resistance	R
A37.	Resistivity	ρ
A38.	Reluctance	$\mathcal{R}$
A39.	Reluctivity	ν
A40.	Rotational frequency	n
A41.	Susceptance	B
A42.	Time constant	τ
A43.	Velocity of light	c
A44.	Wavelength	λ

Additional Symbols

In addition to the 44 symbols listed in Table A, the symbols in Table B are recommended.

Table B

B1.	Cross section	A
B2.	Ampere-turns (see note B2)	AT
B3.	Dynamic power gain (see note B3)	D_P
B4.	Gain (response Function)	$g(t)$, $G(s)$, $G(j\omega)$, $G^*(s)$
B5.	Gain (invariant)	K
B6.	Form factor	k_f
B7.	Length	ℓ
B8.	Current density	J
B9.	Saturable reactor (see note B9)	SR
B10.	Loss, energy, e.g., core loss/cycle	W
B11.	Angle at which the core saturates	α

Note B2. Ampere-Turns—AT

AT can be used as a unit of magnetomotive force, but should not be used as a symbol for magnetomotive force for which the symbol $\mathcal{F}$ has been reserved, see A24 Table A.

Paper 59-173, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted August 18, 1958; made available for printing December 2, 1958.

The personnel of the Theory Subcommittee of the AIEE Magnetic Amplifiers Committee are: H. F. Storm, *chairman*, R. C. Barker, Paul P. Biringer, H. C. Bourne, Jr., A. Boyajian, L. A. Finzi, F. J. Friedlaender, H. L. Goldstein, Alexander Kusko, H. W. Lord, B. W. Manteuffel, R. A. Ramey, R. W. Roberts, A. B. Rosenstein, T. G. Wilson.

Note B3. Dynamic Power Gain—D_P

The term "dynamic power gain" has been in use in Europe for quite a number of years to designate the power gain per second. The symbol in use is D. Since the magnetic amplifier also displays a dynamic voltage gain, and perhaps others, the subscript P was added, resulting in the proposed symbol D_P. In this country there is also the term "figure of merit" in use to express the power gain per second; however, the term dynamic power gain is more specific than figure of merit, which may relate to many meritorious relations.

Note B9. Saturable Reactor—SR

The abbreviation *SR* is analogous to such widely accepted abbreviations as *CT* for current transformer or *PT* for potential transformer.

Lower Case Subscripts

The following lower case subscripts are recommended:

Table C

C1.	Average (see note C1)	*a*, 0
C2.	Coercive	*c*
C3.	Effective (see note C3)	*e*
C4.	Intrinsic	*i*
C5.	Maximum (peak)	*m*
C6.	Residual induction	*r*
C7.	Saturation	*s*
C8.	Reverse rectifier (voltage or current)	*r*
C9.	Forward rectifier (voltage or current)	*f*

Note C1. Average Values

There are two basic kinds of average values in magnetic amplifier work. Let us consider first the following:

$$E_0=\frac{1}{T}\int_0^T e(t)dt \qquad (1)$$

E_0 represents the average of the variable $e(t)$ when averaged over the entire period T, such as when determining the average value of the voltage of a rectifier. The term E_0 is identical with the zero-frequency term of a Fourier analysis of the variable $e(t)$, and hence should be designated by a subscript of 0. E_0 should be pronounced "E subzero."

Another type of average is expressed by the following:

$$E_a=\frac{2}{T}\int_0^{T/2} e(t)dt \qquad (2)$$

E_a represents the average of the variable $e(t)$ when averaged over only one-half of the period, such as when determining the flux swing in one core of a single-phase full-wave magnetic amplifier. This average value is called half-period or half-cycle average, and it is a quantity which is of dominating importance in magnetic amplifier work.

The half-period average should be designated by a subscript of a as shown in E_a. However, since the half-period average is so predominant, the continuous usage of a subscript may be considered cumbersome, in which case the option is offered of not using any subscript at all when designating the half-period average value in magnetic amplifier papers. In polyphase magnetic amplifiers, the integration is carried out over a fraction of the period, different from one-half. The designation for this average is for instance one-third-cycle average, or one-sixth-cycle average, etc., and the notation remains E_a or E.

Note C3. Effective Values

A subscript *e* should be used in magnetic amplifier papers to denote effective (root-mean-square) values in order to avoid confusion between average and effective values.

Upper-Case Subscripts

The following upper-case subscripts are recommended:

Table D

D1.	Ampere-turn	*AT*
D2.	Bias	*B*
D3.	Control	*C*
D4.	Copper	Cu
D5.	Voltage	*E*
D6.	Feedback	*F*
D7.	Iron	Fe
D8.	Gate	*G*
D9.	Current	*I*
D10.	Load	*L*
D11.	Output	*O*
D12.	Power	*P*
D13.	Resistance	*R*
D14.	Rectifier	REC
D15.	Excitation	EXC
D16.	Window	*W*

Upper-Case Versus Lower-Case Letters

There is a note to the American Standard Letter Symbols which states that lower-case letters should be used for instantaneous values. Some of the Subcommittee members made strenuous objections against this rule, when it came to b, h, and f for the instantaneous values of flux density, magnetic intensity, and magnetomotive force, at least for the time being. In cases where the instantaneous character of the variable requires emphasis, it should be designated by $B(t)$, $H(t)$, $F(t)$. No objections were raised against using lower case ϕ and ψ for the instantaneous values of magnetic flux, and of magnetic-flux linkage, respectively, and hence lower-case letters should be used for these two quantities.

Reference

1. LETTER SYMBOLS FOR ELECTRICAL QUANTITIES. *ASA Z10.5-1949*, American Standards Association, New York, N. Y., 1949.

◆

Discussion

G. E. Lynn (Westinghouse Electric Corporation, Baltimore, Md.): The work of the Theory Subcommittee in preparing this paper has resulted in a most welcome addition to the lengthening list of standards available in the field of magnetic amplifiers. It is a foregone conclusion that the symbols presented in the paper represent numerous compromises within the Subcommittee and it is probably not to be expected that the results will be equally palatable to all practitioners in the magnetic amplifier field. With an appreciation of these facts, disagreements stemming principally from personal prejudices or preferences should not be submitted for discussion.

It would seem, however, that some specific symbols proposed violate, in some degree, the aims of standardization. It would not appear wise, for example, to standardize a symbol which differs from accepted standards in other fields merely on the basis of expediency or because of established precedent only in magnetic amplifier literature. Particularly, the proposal that the symbols denoting an rms value carry a subscript while those denoting average values need not seems at variance with standard symbols in other electrical fields. A review of recent magnetic amplifier literature indicates that almost all quantities involving average values already carry a subscript and are defined as average values in the accompanying list of symbols. It would appear that if, in a particular paper or book, an author should feel that use of a subscript for average values would work undue hardship, the author is at liberty to use a nonstandard symbol and define it properly himself. This would eliminate possible confusion arising from someone outside the magnetic amplifier field attempting to understand a paper without first consulting the magnetic amplifier standard symbols list.

In addition to the change suggested in the foregoing, some additions to the list of standard symbols would be desirable. For those who prefer a figure of merit based on power gain per cycle to the dynamic power gain defined in the paper, an adequate symbol should be defined. Also, since control winding refers to any winding other than the gate winding, the use of E_c for the signal voltage can lead to confusion. It is suggested that the capital subscript S be added to Table D of the paper to stand for "signal."

Vol.
Pt. 1
195

H. F. Storm: As the title of the paper indicates the symbols listed therein are merely "recommended" for use. This means that an author fully retains his freedom of choosing symbols according to his wishes.

Mr. Lynn's specific proposals will be submitted to the Theory Subcommittee of the AIEE Magnetic Amplifiers Committee, when amendments and additions will be considered. However, it has been the consensus of the Theory Subcommittee to try the present set of symbols for one year, before introducing changes and additions.

The interest displayed by Mr. Lynn is appreciated very much. Even after the official deadline for formal discussions of the recommended symbols has been passed, members of the Institute are encouraged to further discussions of this matter with the chairman of the Theory Subcommittee.

Observation of Transients in the Series-Connected Saturable Reactor With High-Impedance Control Source

H. L. GOLDSTEIN
ASSOCIATE MEMBER AIEE

IN A SERIES-CONNECTED saturable reactor working in the proportional mode with its control winding fed by a direct current from a high impedance source, the gate current as a function of time has a rectangular shape and its amplitude, in ampere turns, is equal to the magnitude of the control current.[1] Full-wave rectification of the gate current yields a direct current closely proportional to the control current, regardless of changes in line voltage, frequency, and load resistance. Hence the name d-c transformer. This relation holds in general even under transient conditions. For instance, the gate current follows changes in the control current instantaneously and a ripple component in the control current is reproduced faithfully in the gate current even when its frequency is higher than that of the power source.[2]

An extension of Storm's analysis, however, will show that the circuit can be temporarily thrown out of its proportional mode by a transient. Transients are introduced in this investigation by short circuiting the load resistor or removing a short circuit from it at a predetermined instant of the cycle. During the transient, the waveforms of the gate current and load voltage are observed on a scope and the phase angle of the gate current in each cycle is recorded.

Theory

Circuit in Proportional Mode

The circuit under investigation is shown in Fig. 1. Two identical coils, each with the ideal magnetic characteristic shown in Fig. 2, have their control windings connected in series to a d-c source and in series with a large impedance Z, which reduces alternating currents to a small magnitude compared to that of the d-c I_C. The gate windings are connected to a source of sinusoidal voltage $e=E_m \sin \omega t$ (supply voltage), in series with the load resistance R. The relative polarities of the windings are indicated by dots. In the proportional mode of operation, always one of the coils is saturated and the other unsaturated. The gate curent i_G has a rectangular waveshape (Fig. 3), and the unsaturated coil, in which control and gate ampere-turns cancel, takes up the difference between supply voltage and load voltage $i_G R$. Between two polarity changes of the gate current $\omega t=\alpha_1$ and $\omega t=\pi+\alpha_2$, the operating point on the magnetic characteristic of the unsaturated coil moves from the knee into the unsaturated region and back again. The over-all flux swing is zero and the two shaded areas in Fig. 3, consequently, are equal.[1,3,4] This lead to the relation

$$\int_{\alpha_1}^{\pi+\alpha_2} (E_m \sin \omega t - i_G R) d\omega t = 0$$

$$\cos \alpha_1 + \cos \alpha_2 = \frac{I_G R}{E_m} (\pi + \alpha_2 - \alpha_1) \quad (1)$$

I_G = amplitude of gate current

Similarly

$$\cos \alpha_2 + \cos \alpha_3 = \frac{I_G R}{E_m} (\pi + \alpha_3 - \alpha_2) \quad (2)$$

This allows to compute the value of the phase angle (The term "phase angle" of the rectangular current is used here to designate the instant when the current changes polarity.) when that of the previous half-cycle is known. In the steady state

$$\alpha_1 = \alpha_2 = \alpha_\infty$$

and

$$\cos \alpha_\infty = \frac{\pi}{2} \frac{I_G R}{E_m} \quad (3)$$

Under steady-state conditions, the smallest value for α_∞ in the proportional mode is 32.5 degrees; in this case the edge of the $i_G R$ wave coincides with the supply voltage wave.[1] The highest value is 90 degrees, for $I_G R = 0$.

From inspection of equation 1 it follows that when $\alpha_1 < \alpha_\infty$, then $\alpha_2 > \alpha_\infty$. This in turn leads to $\alpha_3 < \alpha_\infty$ (from equation 2). By subtracting equation 2 from equation 1 we obtain

$$\cos \alpha_1 - \cos \alpha_3 = \frac{I_G R}{E_m} (2\alpha_2 - \alpha_1 - \alpha_3) > 0 \quad (4)$$

that is $\alpha_3 > \alpha_1$. This shows that the phase angle asymptotically approaches the steady-state value during a transient. The rate at which the return to steady

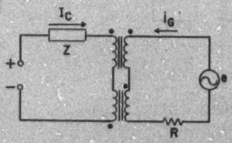


Fig. 1 (left). Series-connected saturable reactor

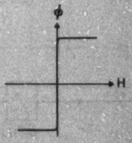


Fig. 2 (right). Ideal magnetic characteristic

Paper 59-175, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted August 22, 1958; made available for printing December 2, 1958.

H. L. Goldstein is with the Bell Telephone Laboratories, Inc., Whippany, N. J.

The author wishes to acknowledge the help of Mr. E. Fischer who assembled the test circuits, assisted in the measurements, and did part of the numerical computations.

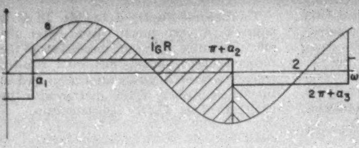


Fig. 3. Load and supply voltages in proportional mode

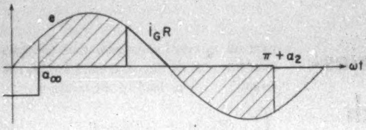


Fig. 4. Current spike after short-circuiting of load

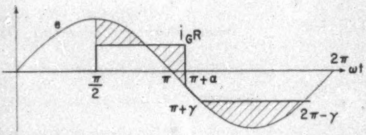


Fig. 5. Rounded edge of current wave after removing short circuit from load

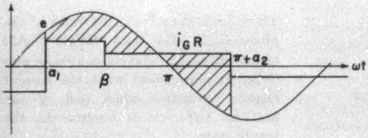


Fig. 6. Immediate transition from one steady state to another (no transient)

Table I. Measured and Computed Values of Phase Angle

n	α_n (°), Measured	α_n (°), Computed
1	34	
2		124.6
3	48.5	48.2
4		115.3
5	57	56.9
6		108.7
7	63.5	63.2
8		103.7
9	68	67.8
10		99.9
11	72	71.4
12		96.9
13	75	74.2
14		94.5
15	77	76.4
16		92.6
17	78.5	78.2
18		91.0
19	80	79.6
20		89.8
21	81	80.7
22		88.9
23	82	81.5
.		
∞	85	

state takes place (amount of damping) is determined for given values of I_G and E_m by the magnitude of R.

Even for low values of R (low damping) deviations of the circuit from the assumed ideal conditions do not introduce any additional damping. This can be seen in the following way. Integrating the gate voltage from α_1 to α_3 leads to

$$\int_{\alpha_1}^{\alpha_3} E_m \sin \omega t \, d\omega t = E_m(\cos \alpha_1 - \cos \alpha_3) = I_G R(2\alpha_2 - \alpha_1 - \alpha_3) \quad (5)$$

The physical meaning of this equation is that the average of the supply voltage taken over a full cycle of the load current is different from zero when there is any damping ($\alpha_1 \neq \alpha_3$). Loosely speaking, the supply voltage has a direct component in this time interval. Since the average voltage across the coils in that time is zero, this direct voltage has to appear across the load, namely as the voltage drop caused by the direct component of the unsymmetrical gate current. Deviations of the circuit from the postulated ideal conditions will result in deviations of the current wave from its ideal rectangular shape. As long as they are small compared to the magnitude of the gate current, their contribution to the damping can be neglected. This explains the good agreement between measured and computed values in Table I.

Deviations from Proportional Mode

Two special cases will illustrate now how the circuit can deviate from its

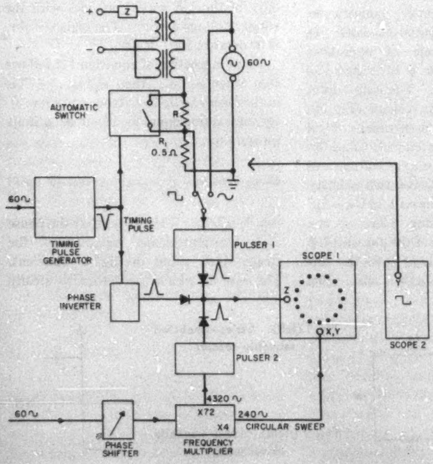

+
−
Z
60∿
AUTOMATIC SWITCH
R
R_1
0.5Ω
60∿
TIMING PULSE GENERATOR
TIMING PULSE
PULSER 1
SCOPE 1
Z
X,Y
PHASE INVERTER
PULSER 2
SCOPE 2
4320∿
X72
X4
240∿
CIRCULAR SWEEP
60∿
PHASE SHIFTER
FREQUENCY MULTIPLIER

Fig. 7 (left). Complete testing setup

Fig. 8 (below). Electronic switch

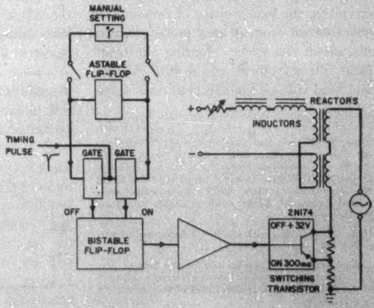

MANUAL SETTING
ASTABLE FLIP-FLOP
TIMING PULSE
GATE
GATE
OFF
ON
BISTABLE FLIP-FLOP
+
−
REACTORS
INDUCTORS
2N174
OFF +32V
ON 300ma
SWITCHING TRANSISTOR

Table II. Data of Coils

Reactors:	
Each Core *51022-2A* (Magnetics, Inc.) Orthonol (50% Nickel-Iron) tape = 2 mil	
Inner diameter, inches	2.0
Outer diameter, inches	3.0
Height, inches	1.0
Winding (1–2) 275 turns, wire no. 17, resistance	= 0.5 ohms
Winding (3–4) 1,650 turns, wire no. 27, resistance	= 34 ohms
Inductors	
Each Core *H95* (Arnold Engineering Company)	
Winding 6,620 turns, wire no. 26	
Inductance = 27 henry at 0.2-ampere direct current	

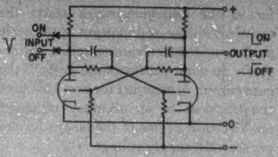


Fig. 9. Bistable flip-flop

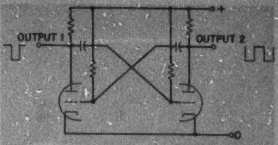


Fig. 11. Astable flip-flop

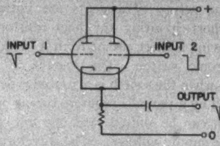


Fig. 10. Electronic gate

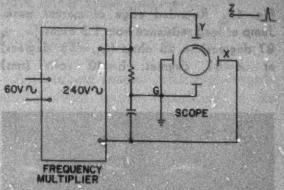


Fig. 12. Circular sweep

proportional mode because of a sudden change in operating conditions.

Jump of Load Voltage from Finite Value to Zero

The voltage across the load is assumed to have some finite value, with the steady-state phase angle α_∞ given by equation 3. At $\omega t=\alpha_\infty$, when the gate current changes polarity, we short circuit the load and thus let I_GR drop to zero; see Fig. 4. The unsaturated coil must now absorb the full gate voltage starting from this moment, that is, earlier than under steady-state conditions with a short-circuited load. Unless the coil is designed with sufficient margin, with respect to the voltage, it will reach saturation before the end of the half-cycle. The full line voltage then will be applied to the gate circuit, resulting in a high peak of the gate current, which under actual conditions is now limited only by winding resistance and saturated inductance of the coils. After the current peak shown in Fig. 4, the current will return to its correct value when the gate voltage reverses polarity at $\omega t=\pi$. The coil is now unsaturated again and its operating point on the magnetic characteristic moves back to the other knee which it reaches at $\omega t=\pi+\alpha_2$. Here the load current changes polarity. When the amplitude of the supply voltage has its highest value consistent with the proportional mode $\alpha_2=\pi/2$. For lower values of gate voltage $\alpha_2>\pi/2$. α_2 will always be smaller than π, since the line voltage applied during a part of the cycle only was able to saturate one coil, which was the very reason for the current peak to appear. This shows that the circuit is now back to its proportional mode.

Jump of Load Voltage from Zero to Full Value

Next we assume I_GR was zero (load resistance short circuited), with the phase shift between line voltage and load current equal to $\pi/2$. At $\omega t=\pi/2$, when the load current changes polarity, we let I_GR jump to its full value; see Fig. 5. Since fewer volt-seconds are absorbed now by the unsaturated coil then under steady-state conditions, the current will reverse polarity earlier, possibly at a time when the supply voltage is not yet high enough to sustain the voltage across the load. The leading edge of the following current wave thus will be rounded off. After that, in most practical cases, the current resumes its rectangular shape again (see the appendix), with its phase gradually approaching the steady-state value as discussed before.

No Transient

The circuit can be switched from one steady-state condition to another without introducing any transient phenomenon by suitably choosing the switching moment $\omega t=\beta$. The conditions for that are (see Fig. 6)

$$\int_{\alpha_1}^{\beta}(E_m\sin\omega t-I_GR_1)d\omega t+\int_{\beta}^{\pi+\alpha_2}(E_m\sin\omega t-I_GR_2)d\omega t=0 \quad (6)$$

In addition, since α_1 and α_2 are steady-state phase angles

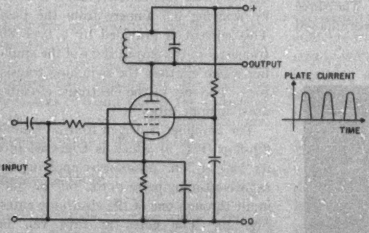


Fig. 13 (above). Frequency multiplier

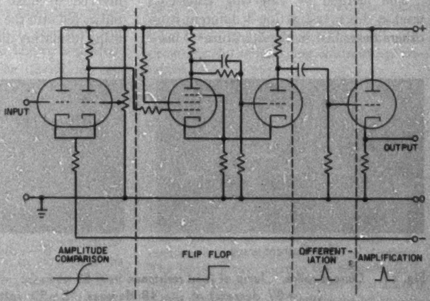


Fig. 14 (right). Pulse generator

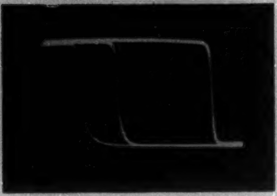


Fig. 15. Rounded edge of current wave. Jump of load resistance from 1.5 ohms (α_∞ = 87 degrees) to 25 ohms (α_∞ = 33 degrees) at ωt = 87 degrees. E = 40 volts (rms)

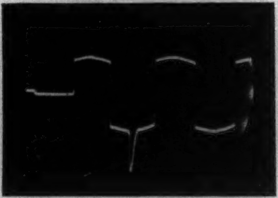


Fig. 16. Voltage spike in negative half-cycle. Jump of load resistance form 1.5 ohms (α_∞ = 87 degrees) to 12 ohms (α_∞ = 73 degrees) at ωt = 87 degrees. E = 40 volts (rms)

$$\cos\alpha_1 = \frac{\pi}{2}\,\frac{I_G R_1}{E_m} \text{ and } \cos\alpha_2 = \frac{\pi}{2}\,\frac{I_G R_2}{E_m} \qquad (7)$$

This allows to compute β. From equations 6 and 7 results

$$E_m(\cos\alpha_1 + \cos\alpha_2) = I_G R_1(\beta - \alpha_1) + I_G R_2(\pi + \alpha_2 - \beta) = \frac{\pi}{2}\,(I_G R_1 + I_G R_2)$$

$$\beta = \frac{R_1(\pi/2 + \alpha_1) - R_2(\pi/2 + \alpha_2)}{R_1 - R_2} \qquad (8)$$

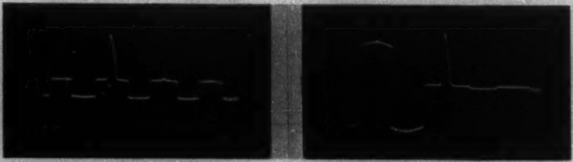
Measurements

The method of observing transients used in this investigation is different from others described in the literature[5-8] insofar as the switching which starts the transient is done at a moment of the cycle predetermined to about one electrical degree. In addition, the phase angle of the gate current is recorded for every cycle of the transient to an accuracy of about one electrical degree of the 60-cps (cycles per seconds) supply voltage.

BLOCK DIAGRAM

Fig. 7 shows a block diagram of the testing setup. Transients are initiated by short-circuiting the load resistor or removing a short circuit from it with an electronic switch. The instant of the cycle when the switch is actuated is determined by a periodic negative timing pulse which can be shifted over one full cycle of the supply voltage. The waveform of the gate current is observed by connecting the *Y*-input of scope *2*, which works with linear sweep, across the small resistor R_1 (= 0.5 ohm) in the gate circuit. In addition, the voltage across the load is observed.

Scope *1*, which measures the phase of the gate current, works with a circular sweep of 240 cps so that one revolution of the beam corresponds to 90 electrical degrees of the 60-cps supply voltage. The instant to be observed is marked on the scope by applying a positive pulse to its *Z* input. The pulse is given out by a pulser circuit at the moment when the voltage applied to its inputs goes through zero in the positive direction. Pulser *1* is connected either to the 0.5-ohm resistor to monitor the gate current, or to the supply voltage to determine the origin of the time scale (zero phase shift). Pulser *2*, which is fed from a voltage of 4,320 cps derived from a frequency multiplier, provides time markers spaced 5 electrical degrees apart. The phase shifter at the input of the frequency multiplier allows to align one time marker with the origin of the time scale. To determine the switching instant, the timing pulse is applied to the scope after passing it through a phase inverter to make it positive. The three pulses fed into the scope are superimposed using insulating diodes.



A—Current wave

B—Voltage wave

Fig. 17. Current spike. Jump of load resistance from 25 ohms (α_∞ = 48 degrees) to 1.5 ohms (α_∞ = 87 degrees) at ωt = 48 degrees. E = 50 volts (rms)

Fig. 18. Transient of phase angle. Jump of load resistance from 19 ohms (α_∞ = 34 degrees) to 2 ohms (α_∞ = 85 degrees) at ωt = 34 degrees. E = 30 volts (rms)

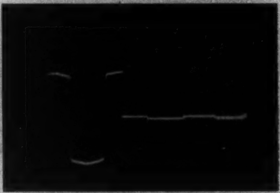


Fig. 19. No transient. Jump of load resistance from 25 ohms (α_∞ = 47 degrees) to 1.5 ohms (α_∞ = 87 degrees) at ωt = 133 degrees. E = 50 volts (rms)

DETAILS OF CIRCUITRY

More details of the circuit under test and the electronic switch are shown in Fig. 8. The data of the reactors and inductors are given in Table II. The a-c component in the control current is about 6 ma (milliamperes) 3%, the exciting current of the reactors in the unsaturated region about 10 ma (less than 1% of the gate current). With a maximum load resistance of 25 ohms, and a gate current amplitude of 1.2 amperes, the maximum amplitude of the load voltage is 30 volts. The load resistor is short-circuited with a *2N174* transistor operated as a switch.[9,10] The latter is kept open, "off" by applying +32 volts between base and emitter, and it is closed by drawing 0.3 ampere from the base. This current is provided by a type *H6* transistor in the output stage of the amplifier which controls the switching transistor. The rest of the electronic circuitry uses vacuum tubes.[11]

The amplifier is driven from a bistable flip-flop (Fig. 9) which is triggered into its "on" or "off" position by applying the negative timing pulse to its "on" or "off" input through one of the electronic gates, Fig. 10. The gates in turn can be controlled in two ways. For visual observations on the scope where it is desired to repeat a transient over and over again, an astable flip-flop (Fig. 11) alternates in

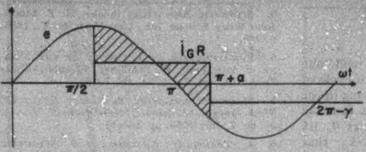


Fig. 20. Rectangular current wave after removing short circuit from load

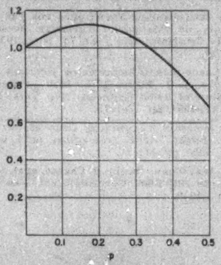


Fig. 21. Ratio of volt-time area absorbed by coil in negative half-cycle of Fig. 20, to highest value occurring in steady state

closing one or the other gate at a low rate. For taking pictures, a single transient is started with a manually operated switch which overrides the astable flip-flop. The timing pulse is taken out of a dynamic voltage tracer made by Hamner Electronics Corporation, Princeton, New Jersey. The delay between timing pulse and closing or opening of the electronic switch (3/4 or 2 electrical degrees, respectively), was considered when setting the timing pulse in the measurements.

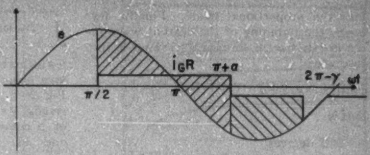
The circular sweep is obtained by applying two voltages 90 degrees out of phase to the horizontal and vertical inputs, respectively, of scope *1*, see Fig. 12. The 240 cps for the circular sweep are derived from a frequency multiplier (Fig. 13) which uses an overdriven penthode and a resonant circuit tuned to the desired harmonic of the distorted plate current. Further multiplication of the 240 cps by factors of 6 and 3 gives the voltage of 4,320 cps required for the time markers. A pulser circuit is shown in Fig. 14.

Measurements

Jump of Load Voltage from Zero to Full Value

Fig. 15 is a picture of the current wave during the transient shown in Fig. 5. Because of winding resistance of the coils and the current sensing resistor of 0.5 ohm, there is still a finite load resistance of about 1.5 ohms left in the load circuit when the transistor is short-circuited. At the right of the picture is the trace before the transient ($\alpha=87$ degrees), in the middle the steady-state condition after switching ($\alpha=33$ degrees) and to the left at $\alpha=16$ degrees the trace immediately after the switching, with the edge rounded off as predicted by theory.

Fig. 22 (right). Voltage spike in neagtive half-cycle



The load voltage spike occurring in the negative half-cycle under the conditions described in the appendix is shown in Fig. 16.

Jump of Load Voltage from Finite Value to Zero

The current spike resulting when the load is short-circuited is shown in Fig. 17(A) (compare Fig. 4). The load voltage (Fig. 17(B)) which is very low in the short-circuited condition, has also a spike because the transistor is not saturated enough to sustain a very-high-load current. Starting with the next half-cycle the current wave essentially is back to its rectangular shape.

Transient of Phase Angle in the Proportional Mode

The change of the phase angle was measured during a transient when the circuit stayed in its proportional mode. The operating conditions are given in Fig. 18 which shows a picture of the scope display. Column 1 of Table I lists the measured values for α, as taken from Fig. 18. The zero point of the time scale is indicated on the photo by a white line. In the second column of the table are values computed with equation 1, assuming the steady state values of α before and after switching as given. The deviations are smaller than 1 degree and therefore within the accuracy of measurement.

No Transient

Under the same circuit conditions as used to obtain Fig. 17, the load was short-circuited at such a moment that no transient resulted at all. The value found experimentally for the proper switching instant, $\omega t=133$ degrees, is very close to the one predicted by theory. (133.5 degrees, from data in Fig. 19 using equation 8.) Fig. 19 shows the load voltage wave under these conditions.

Conclusion

A saturable reactor with high-impedance control source, operating in its proportional mode where gate ampere-turns equal control ampere-turns, can be thrown out of this mode through a transient caused by a sudden change in operating conditions. In the experiments described here, transients are introduced by short-circuiting the load resistance or removing a short circuit from it. The deviation from the proportional behavior which results in a rounding of the normally rectangular gate current wave or a current and voltage spike, occurs only during one cycle after start of the transient. After that the circuit returns to its proportional mode. The phase of the gate current after a transient in general assumes its corresponding new value only gradually with the amount of damping (equals rate of return to steady state) determined by the magnitude of the load resistance in the new condition. By suitably choosing the switching instant, the circuit can be moved immediately from one steady state to another without any transient. Experimental and theoretical results agree within the accuracy of measurement.

Appendix

The course of the transient shown in Fig. 5 depends on the magnitude of the quantity $p=I_GR/E_m$. The maximum value possible for p in the proportional mode is 0.536 ($=\sin 32.5$ degrees); the minimum 0. From Fig. 5 the following relations are obtained

$$\int_{\pi/2}^{\pi+\alpha}(E_m \sin \omega t-I_GR)d\omega t=0$$

$$\cos \alpha=p(\pi/2+\alpha) \tag{9}$$

$$\sin \gamma=p \tag{10}$$

For $p=0.536$, $\alpha=13.7$ degrees and $\gamma=32.5$ degrees; the latter is the steady-state phase angle for the new condition. When p decreases α will increase and γ decrease. The latter means that in the negative half-cycle the unsaturated coil will now absorb more voltage seconds then required to sustain the proportional mode. Finally, for $\alpha=\gamma$, the rounding of the current wave in the negative cycle disappears. That is the case when

$$\cot\text{an}\ \alpha=\pi/2+\alpha$$

that is

$$\alpha=26.3 \text{ degrees}$$

and

$$p=0.443$$

A further decrease in p produces the current waveform shown in Fig. 20. The volt-seconds absorbed by the unsaturated coil in the negative half-cycle now are in proportion to

$$\int_{\alpha}^{\pi-\gamma}(E_m \sin \omega t-I_GR)d\omega t$$

$$=E_m[\cos\alpha+\cos\gamma-p(\pi-\alpha-\gamma)]=E_mf(p) \quad (11)$$

The highest volt-time area one coil must absorb in the steady state is in proportion to

$$\int_{\pi/2}^{\pi}E_m \sin \omega t\, d\omega t=E_m$$

$f(p)$, therefore, is the ratio of the volt-time area absorbed in the negative half-cycle of Fig. 20 to the highest value occurring in the steady state. Fig. 21 shows f as function of p computed by using equations 9, 10, and 11. For $0<p<0.33$, $f(p)>1$; that means in this range the unsaturated coil may reach the other knee of its magnetic characteristic so that the full gate voltage appears across the load; (see Fig. 22). However, a margin of about 12% in the voltage absorbing capacity of the coils would be sufficient to prevent this phenomenon.

References

1. Series-Connected Saturable Reactor with Control Source of Comparatively High Impedance, **H. F. Storm.** *AIEE Transactions*, pt. II, vol. 69, 1950, pp. 1299–1309.

2. A Simple Direct Current Transformer with True Instrument Transformer Properties, **W. Kramer.** *Elektrotechnische Zeitschrift*, Wuppertal-Elberfeld, Germany, vol. 58, 1937, pp. 1309–13.

3. Magnetic Amplifiers (book), **H. F. Storm.** John Wiley & Sons, Inc., New York, N. Y., 1955, p. 168.

4. Fundamentals of a Theory of the Magnetic Amplifier, **W. Schilling.** *Elektrotechnische Zeitschrift*, vol. 71, 1950, pp. 7–13.

5. Magnetic Amplifier Transient Analyzer, **W. A. Geyger.** *Electronics*, New York, N. Y., vol. 26, no. 7, July 1953, pp. 189–91.

6. A Transient Analyzer for Magnetic Amplifiers, **E. J. Smith.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 72, Sept. 1953, pp. 461–65.

7. The Cyclic Integrator—A Device for Measuring the Frequency Response of Magnetic Amplifiers, **T. Dunnegan, Jr., J. D. Harnden, Jr.** *Ibid.*, vol. 73, Sept. 1954, pp. 358–65.

8. Oscillographic Techniques for the Evaluation of Magnetic Amplifier Response, **D. L. Critchlow.** *Ibid.*, vol. 74, Nov. 1955, pp. 607–10.

9. Large-Signal Behavior of Junction Transistors, **J. J. Ebers, J. L. Moll.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 42, 1954, pp. 1761–72.

10. Junction Transistors Used as Switches, **R. L. Bright.** *AIEE Transactions*, pt. I, vol. 74, Mar. 1955, pp. 111–21.

11. Waveforms (book), **B. Chance, et al, editors.** McGraw-Hill Book Company, Inc., New York, N. Y., 1949.

Graphical Analysis of Full-Wave Magnetic Amplifier Control Characteristics Affected by Control-Circuit Resistance

K. MURAKAMI
NONMEMBER AIEE

T. KIKUCHI
ASSOCIATE MEMBER AIEE

SEVERAL PAPERS have been published on the correlation between the control characteristics of magnetic amplifiers and the magnetization characteristics of magnetic cores.[1–5] However, the approaches hitherto inaugurated on this subject of great importance have failed in elucidating the problem concerning the effect of control-circuit resistance (CCR) on the control characteristics of full-wave magnetic amplifiers, because of the lack of an appropriate understanding of the dynamic loop of the reactor core. In the case of the highly oriented magnetic core with rectangular hysteresis loop, the control-magnetization curve (CMC)[6,7] coincides very well with the flux-control characteristics produced by direct-voltage reset.[8,9] This experimental fact enables us to interpret CMC as the relation between the rate of flux change and the instantaneous magnetizing current.

By applying this interpretation and using the method of graphical analysis, the authors have succeeded in solving this problem. The results obtained by the analysis present a very good coincidence with the experimental data.

Experimental Study

As a preliminary to the graphical analysis of the problem, flux-control characteristics of the full-wave magnetic amplifier are measured experimentally at various values of the CCR. A flux-control characteristic is a relation between the average control current and the average induced electromotive force (emf) which is proportional to the amount of reset flux. In order to avoid the effects of a semiconductor rectifier in the gate circuits, and to realize the ideal rectifier condition, mechanical vibrating synchronous contacts are utilized.[8,9] For the reactor core, matched cores of a grain-oriented 50% iron and nickel magnetic material with an almost rectangular hysteresis loop are used.

The experimental circuit is shown in Fig. 1, where the tertiary winding, N_v, and an average sensitive d-c voltmeter together with the synchronous vibrating contacts, C_2, are adopted to measure the amount of flux reset, $\Delta\phi$, during the re-

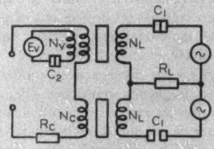


Fig. 1. Experimental circuit

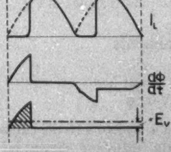


Fig. 2. Current and voltage waveshapes

Paper 59-169, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 15, 1958; made available for printing December 2, 1958.

K. Murakami and T. Kikuchi are with Tohoku University, Sendai, Japan.

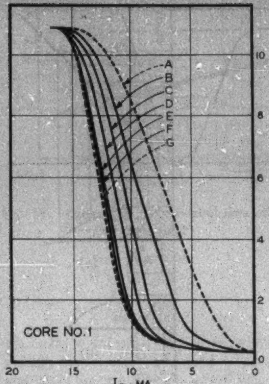


Fig. 3. Flux control characteristics of full-wave magnetic amplifier. R_c equals the following

A—≅0
B—9 ohms
C—34 ohms
D—100 ohms
E—300 ohms
F—1,000 ohms
G—CMC

setting half-cycle of the core.[7] Fig. 2 explains the relation between the output current, I_L, the rate of flux change of the core, $\Delta\phi/dt$, and the rectified voltage applied to the d-c voltmeter, of which the average value is E_g. Then the following relation is satisfied:

$$E_g = fN_g\Delta\phi \tag{1}$$

where f is the frequency of a-c supply voltage. The measured flux-control characteristics are shown in Fig. 3, taking various values of the CCR as a parameter. Corresponding to Fig. 3, the control characteristics of the full-wave magnetic amplifier are shown in Fig. 4. If the average value of output voltage is represented by V_L, its maximum value by V_{Lm}, the number of output winding by N_L, its resistance by R_G, and the load resistance by R_L, then there exists the following relation:

$$V_L = (V_{Lm} - 2fN_L\Delta\phi)R_L/(R_L + R_G) \tag{2}$$

In Fig. 3, the flux-control characteristics for the case of negligibly small CCR cannot be obtained by direct measurement because of the finite value of control-winding resistance. To find it, therefore, the graphical method in Fig. 5 is used. In the figure, the relation between the CCR and the average control current is plotted,

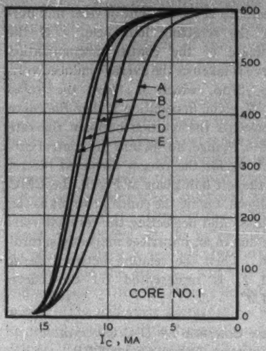


Fig. 4. Control characteristics of full-wave magnetic amplifier. Curves A through E correspond to curves B through F of Fig. 3

taking several values of flux resetting as a parameter. By extrapolating the curves to the side of zero CCR, their intersecting points with the ordinate are obtained which show the control current for the extremely low CCR. Curve *A* in Fig. 3 is thus obtained. For very large CCR, curve *G* of Fig. 3 approaches the control magnetization curve.

Fig. 6 shows the waveforms of induced voltage in magnetic-reactor cores for several values of CCR, their flux resetting level being kept constant.

From the induced voltage waveforms in Fig. 6, it is concluded that the core is reset by almost constant rectangular voltage in the case high CCR. When its resistance is low, the core is reset by the feedback voltage from the other core. For the intermediate value of CCR, both the d-c control voltage and the feedback voltage are effective and the resultant impressed voltage of the reactor winding can be assumed to take the waveform shown in Fig. 7, of which S_1 is due to the d-c control voltage and S_2 to the feedback voltage from the other reactor core.

Graphical Analysis

In this section, starting with the CMC, the method to obtain graphically the flux-control characteristics will be described, and its coincidence with the experimental result demonstrated.

INTERPRETATION OF CMC

As was pointed out elsewhere,[9,10] in the case of grain-oriented 50% iron and nickel magnetic core, the CMC agrees very well with the flux-control characteristics of the direct-voltage reset. In the latter, the group of minor hysteresis loops is observed experimentally to be almost rectangular as shown in Fig. 8. Here, the width of minor loop is seen to be a function of little more than the rate of flux change except in the portion near the negative saturation level. Therefore, it can be interpreted that the width of the dynamic hysteresis loop, due to the reset voltage of arbitrary waveform, is a function of the instantaneous rate of flux change. From this interpretation and the fact that the flux-control characteristics produced by direct-voltage reset coincides with the CMCs, these curves can be

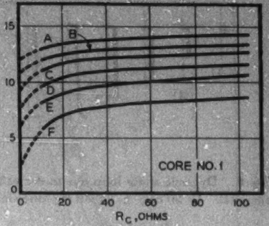


Fig. 5. Extrapolation method to obtain control current for zero control-circuit resistance. $2f\Delta\phi$ equals the following mv per turn

A—10
B—9
C—8
D—8
E—4
F—1

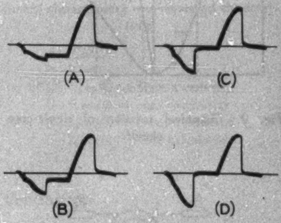


Fig. 6. Induced voltage waveshape affected by control-circuit resistance. $2f\Delta\phi = 6$ mv per turn, R_c equals the following

A—1,000 ohms
B—300 ohms
C—100 ohms
D—10.44 ohms

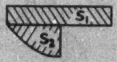


Fig. 7. Resultant reset voltage shape

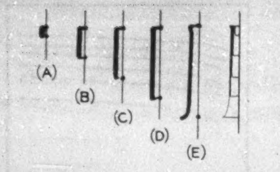


Fig. 8. Dynamic minor loop due to d-c voltage reset. $f=50$ cps, E_r equals the following

A—0.1 volt
B—0.3 volt
C—0.5 volt
D—0.7 volt
E—0.9 volt

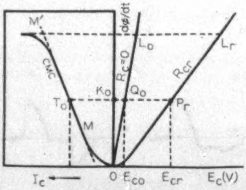


Fig. 9. Graphical solution of single-core circuit

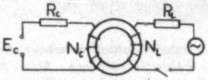


Fig. 10. Single-core circuit under d-c voltage reset

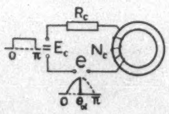


Fig. 11. Equivalent circuit of resetting half-cycle

interpreted as the relation between the instantaneous rate of flux change, $d\phi/dt$, and the control current, I_c. For the larger value of $d\phi/dt$, the straight portion of each curve should be extended linearly. This is due to the experimental fact that the increase of the width of minor loop in the case of direct-voltage reset is proportional to the increase of the rate of flux change for larger values of $d\phi/dt$ (see page 186 of reference 8).

In the case of direct-voltage reset

$$d\phi/dt=\Delta\phi/T/2=2f\Delta\phi \tag{3}$$

where $\Delta\phi$ is the amount of reset flux during the resetting half-cycle. If the ordinate of the control magnetization curve is taken as the average induced voltage, $2f\Delta\phi$, which is twice the value E_r/N_c given in equation 1, then the curve represents the relation between the rate of flux change, $d\phi/dt$, and the control current, I_c, for the reason discussed before. In the left half plane of Fig. 9, the CMC is shown, taking the control current as the abscissa and neglecting the residual voltage due to an imperfect magnetic saturation. Then the relation between $d\phi/dt$ and I_c is represented by the curve OMM'.

Flux Control by D-C Control Voltage Affected by CCR

In order to calculate the flux-control characteristics of the full-wave magnetic amplifier, it is convenient to know, as a preparation, the relation between the voltage impressed on the control winding and the rate of flux change affected by the finite CCR.

Consider the case of a single-reactor circuit shown in Fig. 10. At first, in the case of negligible CCR, the following relation is satisfied between the control voltage and the rate of flux change:

$$E_c=N_cd\phi/dt \tag{4}$$

This relation is represented by a straight line, OL_0 in Fig. 9, taking an appropriate unit of control voltage for the right half of the abscissa. If E_c is constant during the resetting half-cycle, then

$$E_c=2fN_c\Delta\phi \tag{4A}$$

where $\Delta\phi$ is the amount of reset flux during the half-cycle.

Next, taking into account the CCR, R_{cr}, the following relation is satisfied between the control voltage, E_{cr}, and the rate of flux change:

$$E_{cr}=N_cd\phi/dt+R_{cr}I_c \tag{5}$$

where the value of the control current is determined from the left half of Fig. 9 for a specified value of $d\phi/dt$. Then, taking an arbitrary point, K_0, on the ordinate, a straight line $T_0K_0P_r$ is drawn parallel to the abscissa. Let Q_0 be the point of intersection with the line OL_0, and T_0 the intersection with the CMC. Then, for the specified rate of flux change represented by the length of ordinate OK_0, the following relations are satisfied:

$$K_0Q_0=N_cd\phi/dt \tag{6}$$

$$T_0K_0=I_c \tag{7}$$

Using this value of I_c, the voltage on the CCR, R_{cr}, is calculated, which corresponds to the length Q_0P_r; that is

$$Q_0P_r=R_{cr}I_c \tag{8}$$

The point, P_r, is thus determined on the extension of the straight line, T_0Q_0.

If

$$K_0P_r=K_0Q_0+Q_0P_r \tag{9}$$

then, from equations 5, 6, 8, and 9,

$$E_{cr}=K_0P_r \tag{10}$$

Repeating the procedure, the curve OL_r is obtained which represents the relation between the rate of flux change and the voltage applied to the control winding, of which the resistance is R_{cr}.

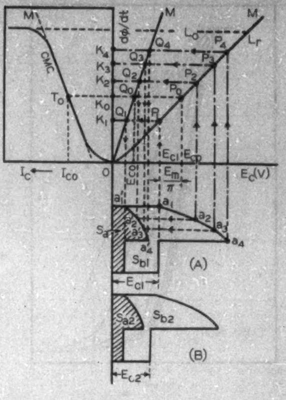


Fig. 12. Graphical analysis of flux-control characteristics

Graphical Analysis of Flux-Control Characteristics

Fig. 11 shows an equivalent circuit of a reactor core of the full-wave amplifier in its resetting half-cycle, where E_c is the d-c control voltage and e is the feedback voltage due to the induced emf of the other core in its gating period. The firing angle, $\theta\alpha$, of the reactor is given by the following relation:

$$\Delta\phi=(1-\cos\theta_\alpha)E_m/2\pi fN_c \tag{11}$$

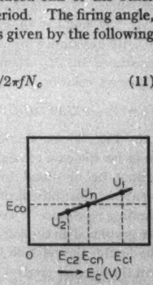


Fig. 13. Method of obtaining corrected control-voltage value

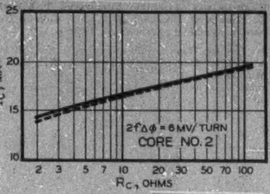


Fig. 14. Comparison of calculated data with the experiment

—— Measured
- - - Calculated

where E_m is the crest value of a-c supply voltage seen from the control winding, and $\Delta\phi$ is the resetting amount of flux of both cores in the case of the steady state.

An example of the graphical analysis to obtain the magnitude of control voltage for a specified amount of reset flux is shown in Fig. 12. In this particular example the steady-state firing angle, θ_α, is adjusted to $\pi/2$ radians. Then the amount of reset flux $\Delta\phi_r$, is given by the following equation:

$$\Delta\phi_r = (E_m/\pi)(1/2fN_c) \tag{12}$$

The problem is to find the correct value of control voltage which gives rise to the given amount of reset flux, $\Delta\phi_r$.

The average value of voltage absorbed in the reactor core during the resetting half-cycle is

$$\begin{aligned} E_{c2} &= 2fN_c\Delta\phi_r \\ &= E_m/\pi \end{aligned} \tag{13}$$

Next, by the method already explained in Fig. 9, the voltage E_{cp} which satisfies the following relation

$$E_{c0} + I_{c0}R_{cr} = E_{cp} \tag{14}$$

is calculated. The average value of volt-seconds which are fed back from the other core is

$$\bar{e} = 2fN_c\Delta\phi_r = E_m/\pi \tag{15}$$

Subtract $\bar{e}$ from E_{cp},

$$E_{cp} - \bar{e} = E_{c1} \tag{16}$$

Then the value of E_{c1} already obtained can be assumed to be the first approximate value of control voltage, which is being sought. Using E_{c1} and the feedback voltage e, the waveform of control-circuit voltage, $E_{c1}+e$, is drawn at A in Fig. 12.

At every instant in a resetting half-cycle, a portion of the voltage is consumed on the CCR and the remainder is absorbed in the reactor core. The graphical procedures to determine these two portions will be self-evident when compared with the case in Fig. 9. Thus, in the voltage waveform A of Fig. 12, the hatched area S_{a1}, which is the volt-seconds absorbed in the core, and the area S_{b1}, which is the volt-seconds on the CCR, can be determined. The half-cycle average of S_{a1} will be found to be a little larger than the voltage E_{c0} given by equation 13, due to the nonlinearity of the curve OL_r.

Next, choose a value of control voltage, E_{c2}, which is a little smaller than E_{c1} and draw the waveform B in Fig. 12, and repeat a similar procedure. The time average of the volt-seconds consumed in the core, S_{a2}, might be smaller than E_{c0}. The correct value of control voltage can be obtained by the interpolation method shown in Fig. 13. In this figure, the abscissa is the d-c control voltage and the ordinate is the time average of the volt-seconds consumed in the core. The point U_1 corresponds to E_{c1} and U_2 to E_{c2}. On the line U_1U_2, choose a point U_n which gives the value of ordinate equal to E_{c0}. Then its abscissa E_{cn} is the correct value of control voltage sought.

The control current is calculated easily by the equation

$$I_{cn} = E_{cn}/R_{cr} \tag{17}$$

Comparison of the Results of Analysis with Experimental Data

For various values of the CCR, the control current is calculated at the same constant flux-reset level which gives the firing angle, $\theta_a = \pi/1$. In Fig. 14 the values of control current, obtained both experimentally and by calculation, are plotted by taking the values of CCR on the abscissa. The good coincidence of the values shows the validity of the presented method.

The whole range of flux-control characteristics is calculated by procedures similar to that shown in Fig. 12, for two values of CCR, i.e., 10.44 ohms and 40.44 ohms. This is compared with the experimental results as shown in Fig. 15, which shows a good agreement of both data.

The data on magnetic cores used in the experiment are shown in Table I.

Table I. Data on Magnetic Cores

Core	Mean Diameter, Millimeters	Weight, Grams	Tape Thickness, Millimeters	Control Winding, Turns
Number 1	45	43.0	0.025	100
Number 2	30	35.7	0.05	50

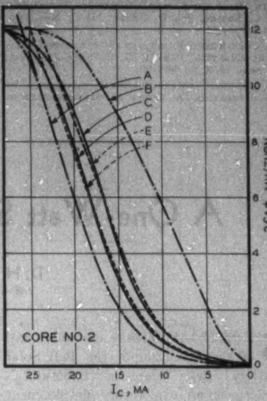


Fig. 15. Comparison of calculated flux-control characteristics with measured characteristics

A—CMC
B—$R_c \cong 0$
C—R_c = 10.44 ohms, measured
D—R_c = 40.44 ohms, measured
E—R_c = 10.44 ohms, calculated
F—R_c = 40.44 ohms, calculated

Conclusions

The effect of CCR on the flux-control characteristics of the full-wave magnetic amplifier is successfully elucidated under the ideal rectifier condition and for the magnetic cores with an almost rectangular hysteresis loop. The correlation between the control characteristics of the full-wave amplifier and its flux-control characteristics is comparatively simple. Therefore, the paper is a contribution to a deeper understanding of the mechanism of the full-wave magnetic amplifier.

Though this paper treats only the case of a steady-state condition, the transient phenomena can be discussed from the same viewpoint.

References

1. THE INFLUENCE OF MAGNETIC AMPLIFIER CIRCUITRY UPON THE OPERATING HYSTERESIS LOOP, Harold W. Lord. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 72, 1953 (Jan. 1954 section), pp. 721–28.

2. THEORY OF MAGNETIC AMPLIFIERS WITH SQUARE-LOOP CORE MATERIALS, H. F. Storm. *Ibid.*, Nov., pp. 629–40.

3. ANALYSIS OF THE FULL-WAVE MAGNETIC AMPLIFIER CIRCUITS CONSIDERING THE CHANGE OF THE WIDTH OF THE DYNAMIC HYSTERESIS LOOP, Tadashi Kikuchi. *Ibid.*, vol. 76, July 1957, pp. 241–49.

4. DYNAMIC OPERATION OF MAGNETIC AMPLIFIERS

with Real-Core and Rectifier Functions, **H. C. Bourne, Jr., D. Nitzan.** *Ibid.*, Nov., pp. 646–55.

5. Core-Reset Functions in Magnetic Amplifiers Analysis, Part I, **G. C. Feth.** *Ibid.*, vol. 77, Sept. 1958, pp. 503–09.

6. Magnetic Characteristics Pertinent to the Operation of Cores in Self-Saturating Magnetic Amplifiers, **R. W. Roberts.** *Ibid.*, vol. 73, 1954 (Jan. 1955 section), pp. 682–90.

7. Measurement of Control Magnetization Curve of Magnetic Amplifier Core Using Vibrating Rectifiers, **T. Kikuchi, K. Murakami.** *Electrotechnical Journal of Japan*, Tokyo, Japan, June 1956, pp. 44–46.

8. On the Flux Control of the Saturating Reactor for Magnetic Amplifiers, **T. Kikuchi, K. Murakami.** *Technology Reports*, Tokohu University, Sendai, Japan, vol. 20, 1956, p. 179.

9. On the Input Power of the Half-Wave Magnetic Amplifier Circuit, **T. Kikuchi, K. Murakami.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, Mar. 1957, pp. 10–14.

A One-Watt Solar Power Plant

D. H. SMITH
MEMBER AIEE

SHORTLY AFTER the announcement[1] of the invention of the Bell Solar Battery, Bell Telephone Laboratories undertook an exploratory development program with a twofold purpose. First, device designers wanted to learn how to make large quantities of silicon photodiodes, how to devise reliable connections to the active parts of the diode, and if possible, how to increase the realizable efficiency of conversion of sunlight to electricity. Second, system designers wished to find out how to use the new photodiodes in the most efficient and practical way in power plants for telephone equipment.

Reasonable progress has been made in this exploratory program. Efficient silicon photodiodes with reliable electrical contacts were fabricated in quantity for the exploratory program, and similar diodes are now available on the commercial market.

In addition, the fundamental system design problems have been explored and the solutions to a number of these problems have been obtained during the engineering planning for a recent field experiment.

This paper describes some of the preliminary work that culminated in the field experiment, including exploratory device development and engineering studies. Preliminary tests are discussed, and field test data are summarized.

Exploratory Development

An elementary solar battery, or silicon photodiode, is a small circular wafer cut from a single crystal of negatively-doped silicon. About one ten-thousandth of an inch below the surface of the wafer is a p-n junction, formed by the gaseous diffusion process at a carefully controlled temperature. A cross section of one of wafers is represented in Fig. 1. When a wafer 1 inch in diameter is exposed to sunlight, it intercepts about 1/2 watt of solar power. Theory predicts that about 1/5 of the solar power, or 100 milliwatts, can be made available to a resistance load connected across the diode. Practically however, only 25 to 50 milliwatts were obtainable from most of the diodes available for use in this exploratory development program, and the average power output per diode in the experimental plant was even lower due to paralleling losses.

When the active surface of the silicon photodiode is subjected to various intensities of illumination, the generated emf (electromotive force) varies somewhat. Therefore, if the load contains a source of energy, such as a battery, protection must be provided to prevent discharge of the battery through the photodiodes when the generated voltage falls below the emf of the battery.

DEVICE DEVELOPMENT

The diode fabrication program presented many challenging problems. Discussion of this part of the development can be found in the literature.[2–4] It is sufficient for this paper, however, to state that crystal purification, doping, diffusion, and contact techniques were worked out, which resulted in a relatively high yield of diodes that had very uniform characteristics. A typical diode was about 1 inch in diameter and four-hundredths of an inch thick. The diode potential in sunlight was about 1/2 volt on open circuit, and the short-circuit current was about 100 milliamperes. Because of internal resistance drop, the voltage under load was about 1/3 volt.

System studies, which paralleled the diode fabrication program, resulted in a recommendation for develpment of a power plant to produce 1 watt average and 10 watts peak. This power plant would be installed near Americus, Ga. at the site of the field tests of a new rural carrier telephone system.[5] It was planned that the solar power plant would furnish electricity to a transistorized carrier terminal for demonstration purposes. Since 432 diodes were required for the power plant, the diode development program was expanded to include the packaging of groups of 9 diodes in modular form. About 105 modules were fabricated; 48 for each of two complete power plants, plus 9 spare modules.

Let us examine a typical module, shown in Fig. 2. Nine photodiodes are mounted on a fiber board and are connected in series by means of etched or "printed" wiring. The board is suspended within a transparent Plexiglas case for mechanical protection. The case is filled with silicone oil to reduce loss of light due to refraction, and the case is sealed. When tested in full sunlight, a typical module delivered about 70 milliamperes at 2.5 volts.

The volt-amp (ampere) characteristics of module no. 8 were measured in the

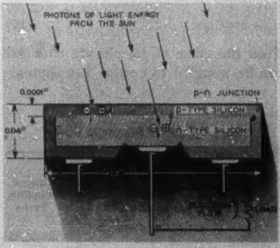


Fig. 1. Cross section of silicon photodiode

Paper **58-186,** recommended by the AIEE Solid State Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted November 4, 1957; made available for printing April 16, 1959.

D. H. Smith is with the Bell Telephone Laboratories, New York, N. Y.

The author wishes to acknowledge the significant technical contributions made by members of several departments at Bell Telephone Laboratories, and the co-operation of the Southern Bell Telephone Company and the Operating and Engineering Department of the American Telephone and Telegraph Company. The author regrets that space does not permit personal acknowledgement of each individual who participated.

Vol.
Pt. I
195

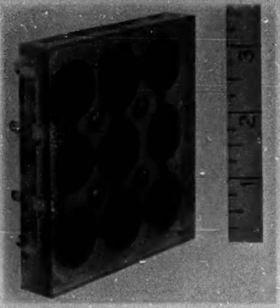


Fig. 2. Typical photo-module

laboratory in artifical (tungsten lamp) illumination, and are plotted in Fig. 3. The curve marked 1.45 for example, was obtained by exposing the module to a radiation intensity approximately equivalent to the average value of full sunlight in the continental United States, 1.45 gram-calories per square cm (centimeter) per minute, or 1,000 watts per square meter. The load on the module was varied from an open circuit (X-axis intercept at 4.7 volts) to a short circuit (Y-axis intercept of 71 milliamperes) and a continuous plot of the volt-amp relationship was obtained. Characteristics were taken at several different intensities; the minimum intensity was about 0.2 gram-calorie per square cm per minute, or about 140 watts per square meter, which is about the intensity of sunlight received on a very dull, cloudy day.

Laboratory tests were also made on individual diodes, and in Fig. 4 the power output and efficiency of a sample diode are plotted against light input. The radiant intensity was increased to a maximum of about 25 times normal sunlight. Output limiting, or saturation, is quite obvious, and there is not much doubt that internal loss is the dominant factor. Reduction of internal losses is an important goal particularly if light-concentrating techniques are to be employed in future exploratory studies.

Circuit Development

The power plant circuit is shown in Fig. 5. The charging source consisted of 48 9-cell photomodules, *M1-M48*, connected to provide a peak output of about 1/2 amp at 22 volts. A silicon diode gate, *CR1*, was used to prevent discharge of the storage battery through the low forward resistance of the photomodules.

Three discharge circuits were connected to the storage battery, *B*1. Fuse *F1* supplied the carrier transmission circuits with a nominal potential of 22 volts through a low-pass *L-C* filter. A bias potential of 16 volts was derived from the 22-volt source through a shunt voltage limiter consisting of a silicon varistor *CR2* operated in its breakdown region, and a ballast resistor *R1*. Fuse *F2* supplied 22 volts to carrier signaling apparatus.

Fuse *F3* supplied operating power to an audio oscillator which provided a 4,100-cycle tone through the low-voltage contact of a voltmeter relay to a remote detector. If the load voltage dropped below a safe value the low-voltage contact opened, causing an interruption in the tone and subsequent operation of the remote detector circuit.

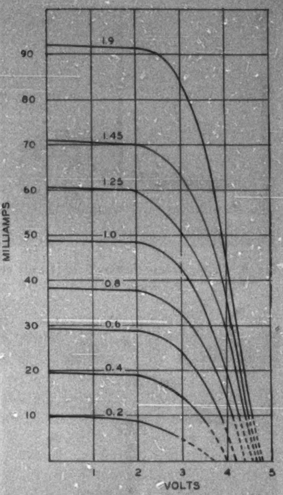


Fig. 3. Volt-amp characteristics of photo-module in artificial light. Numbers on curves refer to $\frac{\text{gm-cal}}{\text{cm}^2\text{-min}}$

System Development

The major system development problems: 1. to devise a weatherproof housing for 48 modules that could be aimed at a particular point in the sky and then locked in position, and 2. to provide means for storing energy for use during the night and during inclement weather. The final power plant design is shown in an over-all view in Fig. 6, a photograph taken at the test site near Americus, Ga.

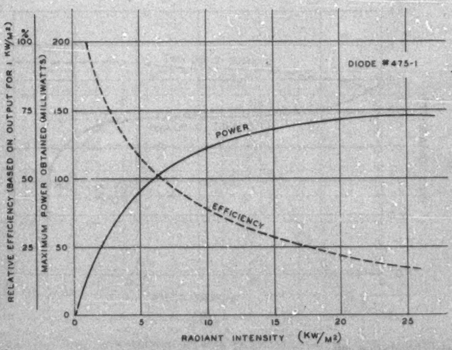


Fig. 4. Power output of a silicon photodiode

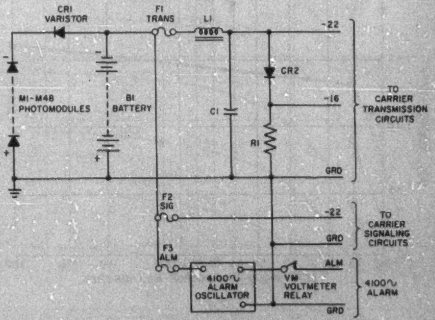


Fig. 5. Solar power plant circuit

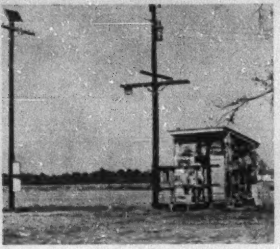


Fig. 6. View of solar battery field test site at Americus, Ga.

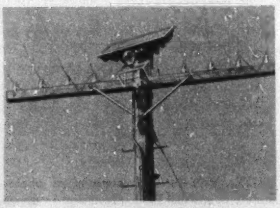


Fig. 7. Closeup of light transducer and reference standard at field test site

At the top of the left-hand telephone pole in Fig. 6 is the light transducer, a glass-front aluminum casting containing the 48 9-cell modules. Near the bottom of the pole are two white, die-cast aluminum boxes. The uppermost box contains the batteries used for energy storage, and a low-voltage alarm circuit. The box nearest the ground contains a type *P1* carrier system terminal. Adjacent is a test shack.

Fig. 7 is a closeup of the light transducer. Noticeable at the top are the stainless steel rods that served to discourage bird roosting and also acted as lightning rods. Just below the light transducer is a bracket upon which is mounted a secondary standard for measurement of radiant intensity. This is a specially selected silicon photodiode, mounted in a sealed glass bulb. The terminals of the secondary standard were connected to a continuous current recorder in the test shack.

Fig. 8(A) is a closeup of the storage battery box with the door open. At the top is a voltmeter relay and an audio oscillator which were part of the low-voltage alarm circuit. Along the hinged side is the filter used to isolate talking and signaling battery, and in the middle is the battery cover, shown open in Fig. 8(B). A total of 96 $1^1/_4$-volt storage cells were used to provide a 10-day reserve. These storage cells, furnished by a commercial vendor, were an experimental nickel-cadimum type which had several unusual features.

1. Each cell was hermetically sealed in a metal and plastic case which prevented the escape of moisture. Thus, the cells did not require the periodic addition of water, or any other routine maintenance. The importance of this feature would vary with the acessibility of the power plant to maintenance personnel.

2. To prevent the buildup of excessive pressure within the sealed container the cells used, in a novel manner, an electrochemical principle[6-8] which involves recombination of the gases formed within the cell during charge and continued overchange.

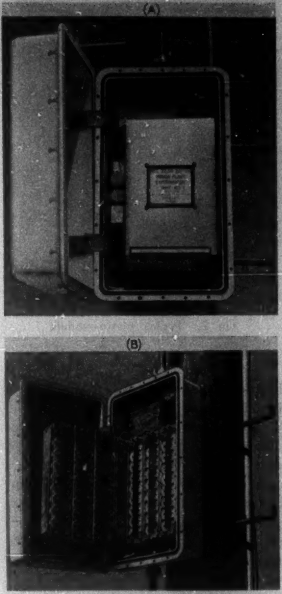


Fig. 8. Closeup of storage battery box

A—Internal cover closed
B—Internal cover open

3. The same mechanism which facilitated the internal recombination of evolved gases also reduced the magnitude of the voltage change associated with a given change in overcharge current, to a value significantly lower than for conventional unsealed storage batteries.

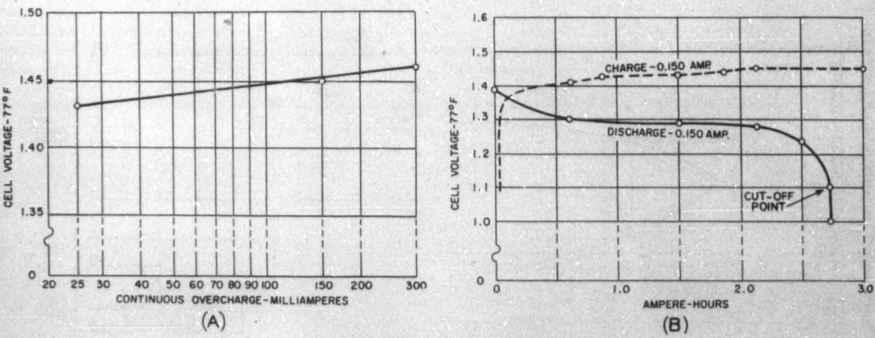


Fig. 9. Sealed nickel-cadmium cell

A—Overcharge characteristics

B—Charge-discharge characteristics

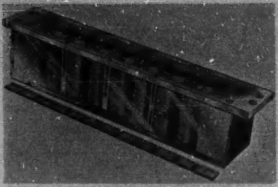
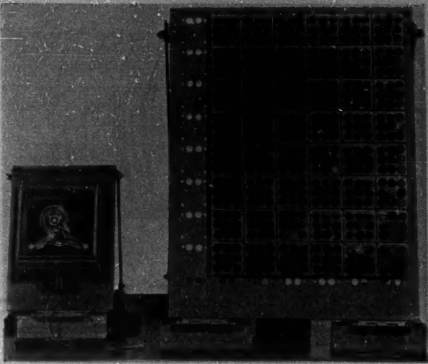


Fig. 10. A 16-cell module of storage battery

Fig. 12. Laboratory test assembly of silicon photomodules



This characteristic is illustrated in Fig. 9(A). The cells could be overcharged continuously at a rate as high as 0.3 amp. This capability of the sealed cells to operate with high charging rates and to take overcharge tended to minimize, through voltage limiting action, the changes in load voltage cause by diurnal changes in the sun's position with respect to the plane of the transducer and the annual changes in the inclination of the plane of the sun's orbit.

The sealed nickel-cadmium cell exhibited the flat charge and discharge characteristics of the nickel-cadmium system, illustrated in Fig. 9(B). Variations in battery voltage from day to night winter to summer, no load to full load, and maximum overcharge to end-of-discharge were so small that an auxiliary voltage regulator was not needed.

Sixteen cells were connected in series to form a 22-volt battery or "storage module," as shown in Fig. 10. Six of these storage modules, rated at about 2.5 amp-hours storage capacity each, were connected in parallel to form a storage battery rated at 22 volts and 15 amp-hours. Fifteen amp-hours was chosen after an investigation of the effect of observed solar radiation in Atlanta, Ga., and in particular, the frequency of extended periods of low radiation which might have resulted in a dangerously low state of charge of the storage battery. Fifteen amp-hours corresponded to a reserve of about 10 days.

On October 4, 1955, the carrier terminal was connected to the experimental power plant, and for the next 6 months all the power required by the terminal was supplied by the sun.

Preliminary Planning and Tests

Before the final design of the power plant could be formulated, it was necessary to determine how much electric energy could be obtained from the sun. Important factors in the planning were the amount of sunlight available at Americus, and the electrical and optical properties of the photodiodes. With these data available, the design effort could proceed.

Light Measurements at Americus

Measurements were made in 1954 of the amount of sunlight at the test site near Americus during the shortest day of the year, December 22. The maximum value on a cloudless day was found to be

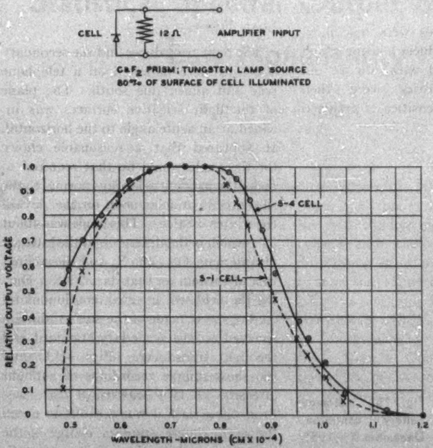


Fig. 11. Spectral response of secondary standard cells

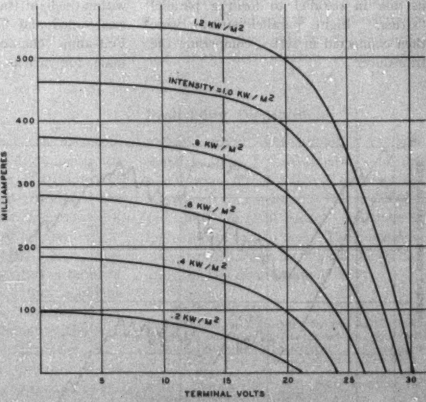


Fig. 13. Volt-amp characteristics of light transducer

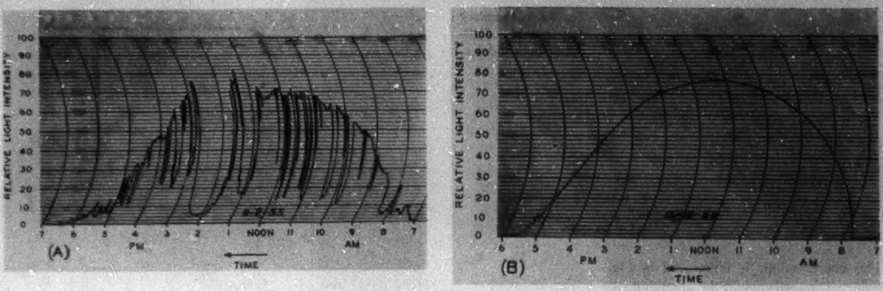


Fig. 14. Chart recordings of sunlight intensity

A—September 2, 1955: mostly sunny, scattered clouds

B—December 31, 1955: full sun, no clouds

0.91 gram-cal/cm²/min (calorie per cm² per minute), or about 633 watts per square meter, on a horizontal surface. Correction of this amount of radiant energy to a plane normal to the sun at its zenith gave 1.6 gram-cal/cm²/min, or about 1,100 watts per square meter which was used as the design value, or reference light.

EFFICIENCY OF SILICON MODULES

Each module received was subjected to a static volt-amp test in various intensities of artificial light. From these data were calculated the optimum load current and the conversion efficiency in reference light. Optimum load currrent varied from 61.5 milliamperes for module No. 10, to 116 milliamperes for module No. 60. The most inefficient module was No. 47 at 5.49% and the most efficient was No. 87 at 7.9%. The data were used to select six modules for connection in parallel to form a parallel "string." Eight parallel strings were then connected in series, comprising the transducer.

COLOR MATCHING

The spectral response of the photodiodes varied from cell to cell as shown in Fig. 11, and to a lesser extent from module to module. The effect of this variation on the plant design was minimized by selecting modules, to be connected in parallel, which had similar spectral response characteristics.

LOAD TESTS

After 48 suitable modules were selected for optimum load current and efficiency, they were assembled on a metal plate, as shown in Fig. 12, and were subjected to a load test, at a fixed voltage in clear sunlight. The results were in accordance with calculations, and it was decided to proceed with final assembly and shakedown tests.

SHAKEDOWN TESTS

The silicon modules were assembled within the light-transducer housing which was sealed and filled with silicone oil. Volt-amp characteristics were then taken for various intensities of artificial light, as shown in Fig. 13. Inspection of Fig. 13 reveals that at a peak radiant intensity of 1.2 kw per square meter, the power plant delivered 1/2 ampere at 20 volts, or 10 watts. This was very close to the design value. The storage batteries, alarm relay, and alarm oscillator were assembled in the storage housing, and the power plant was mounted on the roof of the Bell Telephone Laboratories building in New York, N. Y. The plant was operated into a resistance load for several weeks and performed satisfactorily. The total active silicon area was calculated to be 266.12 square inches and, in full sunlight, the over-all efficiency was found to be 5.8%. The plant was then crated and shipped to Americus, Ga.

Field Tests

LOAD TESTS

The light transducer and the secondary standard were mounted on a telephone pole and aimed due south. The plane of the light sensitive surfaces was inclined at an acute angle to the horizontal. It appeared that a reasonable choice for the angle would be that required to make the incident solar rays normal to the silicon cells at solar noon on the day of the winter solstice. This angle was about 55°, which is the difference of the latitude of Americus, Ga., 32° N, and the inclination of the sun on that day, 23°S. During the field test interval, continuous recordings were made of the sun's radiation received on the secondary standard and the light transducer. Figs 14(A) and (B) show sample recordings of sunlight intensity at two seasons of the year. From these data it was possible to check the approximate state of charge of the battery, including the minimum condition of charge during the period.

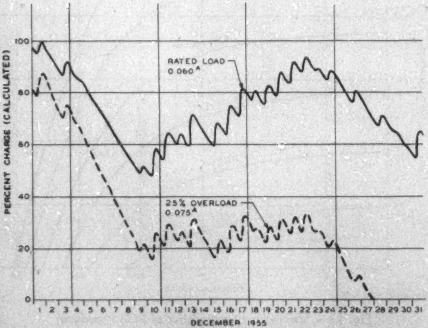


Fig. 15. Storage battery condition, December 1, 1955 through December 31, 1955

The approximate state of charge of the storage battery was calculated, by a simple numerical procedure, on a day-to-day basis during the test. A portion of these calculated data are plotted in Fig. 15, covering the interval during which the lowest state of charge was expected. The computation was based on an assumed value of average load current, and on assumed relationships between charging efficiency and state of charge of the storage battery. On the basis of the calculations, it was concluded that that the battery was half charged on December 10, 1955. It was computed that a load current 25% higher, 75 milliamperes, would have discharged the storage battery completely by December 27, 1955.

Although dirt was deposited on the glass face of the light transducer, natural washing during rainy periods was very effective, and the residual light attenuation, as measured at the Laboratories upon completion of the field tests, was less than 10%. This may not be typical of any other location. No bird deposits were found on the transducer, and no lightning strikes were reported.

The power plant was judged relatively easy to install. Aiming of the transducer for optimum altitude was simplified by a built-in quadrant and plumb-line arrangement. The azimuth setting was made with the help of a compass. A critical inspection of the plant after its return revealed no mechanical damage to either the light transducer or the storage assembly. Although the transducer weighed 125 pounds, including 2 gallons (16 pounds) of silicone oil, and had been exposed to a wide range in ambient temperature, no leaks developed in the gaskets nor was there any evidence of undue mechanical strain.

On the basis of the tests performed in the laboratory and in the field, it was concluded that technical feasibility had been demonstrated for a power plant capable of furnishing about 1 watt.

Economic Considerations

An analysis of the cost of the experimental power plant revealed that it could not compete on an annual charge basis with either of two standard power plants for the type *P1* carrier system. The two standard plants, described in detail elsewhere,[9] consist of a regulated rectifier and an auxiliary storage battery for reserve or, where commercial alternating current is not economically available, a nonrechargeable primary battery of the air-depolarized type.

The principal sources of expense in the solar power plant were the photodiodes and the storage battery. Photodiodes having higher efficiency than those used during the test are now commercially available at about $2.00 per cell, which represents a 10-to-1 price reduction in less than 2 years. Another major price reduction, together with a reduction in the cost of storage elements, could make a solar power plant of the type described competitive with the air-depolarized battery at least for continuous power consumption levels in the order of 1 watt.

References

1. New Silicon P-N Junction Photocell for Converting Solar Radiation into Electrical Power, **D. M. Chapin, C. S. Fuller, G. L. Pearson.** *Journal of Applied Physics*, New York, N. Y., vol. 25, 1954, p. 676; also *Bell Laboratories Record*, New York, N. Y., July 1955; also *U.S. Patent no. 2,780,765*, Feb. 15, 1957.

2. Silicon Solar Energy Converters, **M. B. Prince.** *Journal of Applied Physics*, May 1955.

3. Diffusion of Donor and Acceptor Elements in Silicon, **C. S. Fuller, J. A. Ditzenberger.** *Ibid.*, vol. 27. 1956, p. 544.

4. Radioactive and Photoelectric P-N Junction Power Sources, **W. G. Pfann, W. Van Roosbroeck.** *Ibid.*, vol. 25, 1954, p. 1422.

5. The Type P1 Carrier System (symposium), **R. C. Boyd, D. H. Smith, E. K. Eberhart, F. J. Hallenbeck, E. H. Perkins, J. D. Howard, Jr.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, May 1956, pp. 188–214.

6. Electrolytic Cell and in Particular, Alkaline Cell, **G. Neumann, U. Gottesmann.** *U. S. Patent no. 2,571,927*, Oct. 16, 1951.

7. Gastight Storage Battery and Method of Manufacturing Same, **G. Neumann.** *U. S. Patent no. 2,636,058*, Apr. 21, 1953.

8. Electrolytic Cell and Battery, **R. A. A. Jeannin.** *U. S. Patent no. 2,646,455*, July 21, 1953.

9. See reference 5, p. 191.

Statistical Spectral Output of Power Law Nonlinearity

OTTO J. M. SMITH
MEMBER AIEE

One of the objectives of analyzing a system is to determine if there are periodic oscillations and, if so, what their amplitude and frequency will be; another objective is to determine the magnitudes of output errors and output powers, when the statistical character of the input signal and noises are known. Methods are given for calculating the output power and the error power for nonlinearities describable by power series for Gaussian and for certain other special kinds of input signals.

The objectives of this statistical treatment of nonlinearities are 1. to present a method of describing non-Gaussian random signals, 2. to show the harmonic power spectrum generated by nonlinearities from random signals, 3. to show the calculation of the output power and the error power in a nonlinear system, and 4. to show the calculation of criteria for stability in nonlinear systems. Non-Gaussian signals will be described in terms of the mechanism by which they could be made from a Gaussian source: either power law devices, or discontinuities such as hysteresis, dead zone, saturation, etc.

The harmonic power spectrum at the output of a nonlinearity is related to the squares, cubes, and higher powers of the autocorrelation function. The output has both harmonic power uncorrelated with the input, and fundamental components which are correlated with the input. The cross correlation (CC) between input and output of a nonlinearity is a transference which can be used both for stability tests and for the determination of the useful output.

Probability Distribution

One method of describing a random signal is to plot the probability of the signal, when it is less than a certain voltage, versus that voltage. This is called a probability integral curve. Its range is limited from zero to one. The normal or Gaussian distribution with zero mean has a probability distribution of

---

Paper 59-265, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted December 30, 1957; made available for printing December 17, 1958.

Otto J. M. Smith is with the University of California, Berkeley, Calif.

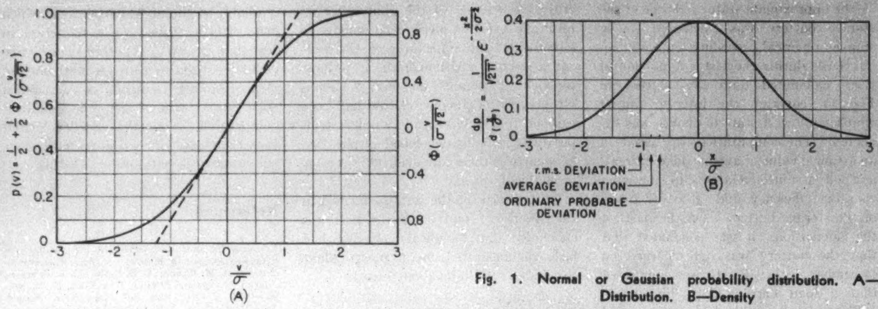


Fig. 1. Normal or Gaussian probability distribution. A—Distribution. B—Density

$$p(v)=\frac{1}{2\pi}\int_{\infty}^{x=v}\epsilon^{-\frac{x^2}{2\sigma^2}}d\frac{x}{\sigma}=\frac{1}{2}+\frac{1}{\sigma\sqrt{2\pi}}\int_0^{x=v}\epsilon^{-\frac{x^2}{2\sigma^2}}dx \qquad (1)$$

The plot of this curve in Fig. 1(A) is called an ogive. If v is plotted versus probability p on probability integral coordinates (Keuffel & Esser type no. 358-23), it is a straight line.

The derivative of the curve of Fig. 1 is the probability density, or differential probability. It is the familiar bell-shaped curve with a mean of zero volts, and inflection points at $\pm\sigma$ volts; see Fig. 1(B).

$$p'=\frac{dp}{dx}=\frac{1}{\sigma\sqrt{2\pi}}\epsilon^{-\frac{x^2}{2\sigma^2}} \qquad (2)$$

This is the probability of a random Gaussian voltage between the values x and $x+dx$. Considering the area under this curve, the ordinary probable deviation is the line which divides each half of the area into halves. This is $\pm 0.674\ \sigma$ voltage deviations which are exceeded half of the time. The center of gravity of each half, or the average deviation, is $0.798\ \sigma$ volts. This value would be delivered by a full-wave linear averaging rectifier. It is the integral of equation 2 times x, over one half of the curve only. The integral over all x is called the first moment, and is zero because the $-x$ terms contribute as much negative weight as do the $+x$ terms positive weight.

The radius of gyration of the area about the mean (zero volts) is the rms voltage σ; this is the second moment. It is the value of the random signal which a dynamometer meter would read. The d-c voltage which a full-wave square-law averaging rectifier would deliver is σ^2.

The mth even moment is

$$M_m=\int_{-\infty}^{+\infty}\frac{x_m}{\sqrt{2\pi}}\epsilon^{-\frac{x^2}{2x^2}}d\left(\frac{x}{\sigma}\right)=1\times3\times5\ldots(m-1)\sigma^m\ (m\text{ even}) \qquad (3)$$

The third moment and all odd moments of a Gaussian signal are zero. If this signal is passed through some full-wave device which reverses the sign of the negative voltages, the computed moments are called the absolute moments.

For this general case, when only half the distribution is used, or when the odd powers of x are all given positive signs, the absolute moments either even or odd, are

$$M_m=2\int_0^{\infty}\frac{x^m}{\sqrt{2\pi}}\epsilon^{-\frac{x^2}{2\sigma^2}}d\left(\frac{x}{\sigma}\right)=\frac{2^{m/2}\sigma^m}{\sqrt{\pi}}\Gamma\left(\frac{m+1}{2}\right) \qquad (4A)$$

When m is odd

$$M_m=\left(\frac{m-1}{2}\right)!\,\frac{2^{m/2}\sigma^m}{\sqrt{\pi}} \qquad (4B)$$

The coefficient in the moment equation will appear many times in subsequent work, and an abbreviation for it is desirable. The coefficient will be defined in terms of a number n, which is one less than the order of the moment; n equals $(m-1)$. The coefficient for the even moments is

$$c_n=c_{m-1}=1\times3\times5\times7\ldots n=\frac{n!}{\left(\frac{n-1}{2}\right)!(n-1)/2} \qquad (5A)$$

When a Γ function is substituted, this series becomes exactly the same as the coefficient in the absolute moment.

$$c_n=c_{m-1}=\frac{2^{(n+1)/2}\Gamma\left(\frac{n}{2}+1\right)}{\sqrt{\pi}}=\frac{2^{m/2}\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{\pi}} \qquad (5B)$$

The gamma function is defined by

$$\Gamma(q)=(q-1)!=\int_0^{r=\infty}r^{q-1}\epsilon^{-r}dr \qquad (5C)$$

This coefficient is valid for all values of n greater than minus 2, and all values of m greater than minus 1; it is not restricted to integer values. With this notation, the absolute moment of any integer or fractional order m of a Gaussian distribution is

$$M_m=c_{m-1}\sigma^m=c_n\sigma^{n+1}=M_{n+1} \qquad (6)$$

High-Order Cross Correlations

The higher moments of the probability distribution of a signal have a meaning which is related to the results obtained from laboratory tests; they are components of the correlation between one power of the signal and a different power. The latter can be measured in practice more easily than the moments, and the transferences of nonlinearities will be in terms of these cross-correlations (CC's), and not the higher moments. The transference will be derived for the general case of two partially correlated signals, from which the special case of a single signal follows directly.

Fig. 2 shows two partially correlated signals. The CC between them is the time average of their product as a function of the time shift (advance) for the second wave.

$$\psi_{ab}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{t=-T}^{t=+T}a(t)b(t+\tau)dt=\mathrm{CC}(a,b) \qquad (7)$$

The CC between the mth power of the voltage a, and the nth power of the voltage b is

$$\psi_{a^mb^n}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{+T}a^m(t)b^n(t+\tau)dt=\mathrm{CC}(a^m,b^n) \qquad (8)$$

The rms value of the signal a will be denoted by σ_a, and for signal b by σ_b; $\psi(0)$ is the cross power in units of volts to the $(m+n)$th power.

The normalized autocorrelation for voltage a is defined by

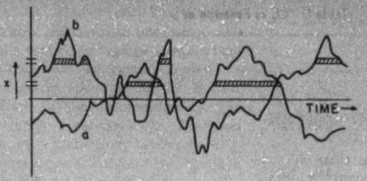


Fig. 2. Time curves of two random functions, illustrating determination of CC as integration over all amplitudes of product of amplitudes, times joint second-order probability density

$$\rho_{aa}(\tau)=\frac{\psi_{aa}(\tau)}{\sigma_a{}^2} \tag{9A}$$

This has a value of unity when τ is zero. The normalized CC is defined by

$$\rho_{ab}(\tau)=\frac{\psi_{ab}(\tau)}{\sigma_a\sigma_b} \tag{9B}$$

This has a value which is always less than one. It equals one, at zero τ, for perfectly correlated signals.

Second-Order Probability Density

The higher-order CC can be derived from probability distributions, instead of from time integrals. The time integral method takes the time average of the value of a times the value of b shifted by a fixed time τ. The density method performs the integration vertically in Fig. 2 instead of horizontally. It takes the product of voltage a times voltage b times the probability that a and b have these particular values simultaneously, and integrates over all values. The simultaneous restriction includes the time shift on b, and therefore includes the τ variable of the CC. More specifically, in Fig. 2 the CC is the integral over all w and x of wx, times the joint probability of a being between w and $w+dw$, while simultaneously b is between x and $x+dx$.

The joint (second-order) probability density for two Gaussian signals is

$$(\partial p_{ab},\partial p_{ab})=\frac{1}{2\pi\sigma_a\sigma_b\sqrt{1-\rho_{ab}{}^2}}\exp-\frac{1}{2(1-\rho_{ab}{}^2)}\left[\left(\frac{w}{\sigma_a}\right)^2-2\times\rho_{ab}\left(\frac{w}{\sigma_a}\right)\left(\frac{x}{\sigma_b}\right)+\left(\frac{x}{\sigma_b}\right)^2\right]\partial w\partial x \tag{10}$$

That this is correct will be shown by the fact that it integrates to the CC; $\rho_{ab}(\tau)$ is the normalized CC

$$\psi_{ab}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}wx\left(\frac{\partial p_{ab}}{\partial w},\frac{\partial p_{ab}}{\partial x}\right)dw\,dx \tag{11}$$

For convenience in integrating, change the variables to y equals w/σ_a, and z equals x/σ_b. Complete the square in the exponent, so that the term enclosed by brackets in equation 10 becomes

$$y^2-2\rho_{ab}\,yz+(\rho_{ab}\,z)^2+z^2(1-\rho_{ab}{}^2) \tag{12}$$

Change variables a second time, letting u^2 equal the squared term divided by the constants in the exponent

$$u=\frac{(y-\rho_{ab}z)}{\sqrt{2}\sqrt{1-\rho_{ab}}} \tag{13}$$

For simplicity, the subscript to ρ can be omitted, and the complete integral in equation 11 then becomes

$$\psi ab=\frac{\sigma_a\sigma_b\sqrt{1-\rho^2}}{\pi}\int_{-\infty}^{+\infty}z\epsilon^{-\frac{z^2}{2}}\times dz\int_{-\infty}^{+\infty}\left(u+\frac{\rho z}{\sqrt{2}\sqrt{1-\rho^2}}\right)\epsilon^{-u^2}du \tag{14}$$

Because of symmetry

$$\int_{-h}^{h}u\epsilon^{-u^2}du=0 \tag{15}$$

The last term in equation 14 is the error integral, or probability integral, defined as

$$\Phi(x)=\frac{2}{\sqrt{\pi}}\int_0^x\epsilon^{-u^2}du \tag{16}$$

Since $\Phi(-x)$ equals $-\Phi(x)$, and $\Phi(\infty)$ is one, the CC becomes

$$\psi_{ab}=\frac{\sigma_a\sigma_b\rho}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}z^2\epsilon^{-\frac{z^2}{2}}dz=\sigma_a\sigma_b\rho_{ab} \tag{17}$$

This reduces to an identity, which demonstrates the method.

Odd-Order CC

The CC between signal a and the nth power of signal b is the integral of the product of w times x^n times the same joint probability that these signals have the values w and x simultaneously.

$$\psi_{ab}n=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}wx^n\left(\frac{\partial p_{ab}}{\partial w},\frac{\partial p_{ab}}{\partial x}\right)dw\,dx \tag{18}$$

$$\psi_{ab}n=\frac{\sigma_a\sigma_b{}^n\sqrt{1-\rho^2}}{\pi}\int_{-\infty}^{+\infty}z^n\epsilon^{-\frac{z^2}{2}}dz\times\int_{-\infty}^{+\infty}\left(u+\frac{\rho z}{\sqrt{2}\sqrt{1-\rho^2}}\right)\epsilon^{-u^2}du \tag{19A}$$

Evaluating the last integral as before yields

$$\psi_{ab}n=\frac{\sigma_a\sigma_b{}^n\rho}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}z^{n+1}\epsilon^{-\frac{z^2}{2}}dz \tag{19B}$$

A considerable similarity between this equation and equation 3 for the higher moments of b is apparent. The integral is zero when n is even. There is correlation only for the odd powers. Evaluating the integral above, yields the odd-order CC

$$\psi_{ab}n=\sigma_a\sigma_b{}^n\rho_{ab}\times1\times3\times5\ldots n=c_n\sigma_a\sigma_b{}^n\rho_{ab} \tag{20A}$$

This is valid only for n odd.

The coefficient is the same as in the $(n+1)$th order moment of signal b. The CC may therefore be written

$$ab^n=\left(\frac{M_{n+1}}{\sigma_b{}^2}\right)\rho_{ab}(\tau)\sigma_a\sigma_b \tag{20B}$$

The coefficient for the nth power CC is the ratio of the next higher moment to the mean square. A higher-order normalized correlation can be defined as

$$\rho_{ab}{}^n=\frac{\psi_{ab}n}{\sigma_a\sigma_b{}^n\times1\times3\times5\ldots n} \tag{21}$$

In this particular case, this reduces to ρ_{ab} when equation 20 is substituted.

Fractional-Order CC

When n is even, b^n is always positive, and there is no correlation between a and b^n. But circuits can be built to deliver $b|b^{n-1}|$ which has the sign of b. These can include fractional exponents and reciprocals of fractional exponents. For this general case, in which b^n has the sign of b, there is an absolute CC, given by changing the integral in equation 19(B) to two times the integral from zero to infinity.

$$\psi_{ab}n=\frac{\sigma_a\sigma_b{}^n}{\sqrt{2\pi}}\left[2^{\frac{n}{2}+1}\Gamma\left(\frac{n}{2}+1\right)\right]=c_n\sigma_a\sigma_b{}^n\rho_{ab} \tag{22}$$

This holds for any n greater than minus 2. When n is minus one, the coefficient c is unity.

Higher-Order CCs

The CC between the mth power of a and the nth power of b is obtained by substituting w^m and x^n in equation 11. Completing the square in the exponent as before yields

$$\psi_{a^mb^n}=\frac{\sigma_a{}^m\sigma_b{}^n(\sqrt{2}\sqrt{1-\rho^2})^m}{\pi\sqrt{2}}\times\int_{-\infty}^{+\infty}z^n\epsilon^{-\frac{z^2}{2}}dz\int_{-\infty}^{+\infty}\times\left[u+\frac{\rho z}{\sqrt{2}\sqrt{1-\rho^2}}\right]^m\epsilon^{-u^2}du \tag{23}$$

The last integral may be rewritten as

$$\int_{-\infty}^{+\infty} f(u)\epsilon^{-u^2}\,du=\int_0^{\infty} f(u)\epsilon^{-u^2}\,du-$$

$$\int_0^{\infty} f(-u)\epsilon^{-(u^2)}\,d(-u)$$

$$=\int_0^{\infty} (f(u)+f(-u))\epsilon^{-u^2}\,du \qquad (24)$$

where $f(u)$ is the binominal series expansion of the bracket to the mth power

$$\left[u+\frac{\rho z}{\sqrt{2}\sqrt{1-\rho^2}}\right]^m$$

in equation 23.

Performing this expansion, adding the two series, and integrating each term, yields

$$\int_{-\infty}^{+\infty} (u+\alpha z)^m \epsilon^{-u_2}\,du$$

$$=2\sqrt{\pi}\left\{\frac{(\alpha z)^m}{2}+\frac{m(m-1)}{2}\,\frac{(\alpha z)^{m-2}}{2^2}+\right.$$

$$\frac{m(m-1)(m-2)(m-3)c_3}{4!}\,\frac{(\alpha z)^{m-4}}{2^3}+$$

$$\ldots+\frac{m!c_{2q-3}}{(m-2q+2)!(2q-2)!}\times$$

$$\frac{(\alpha z)^{m-2q+2}}{2^q}+\ldots+$$

$$\left.\begin{bmatrix} 2^{-\frac{m}{2}-1}\,c_{m-1}, \text{ if } m \text{ is even} \\ \text{or} \\ \alpha z^{2}{}^{-\frac{m+1}{2}}\,c_{m-2}, \text{ if } m \text{ is odd}\end{bmatrix}\right\} \qquad (25)$$

where α is $\rho/\sqrt{2}\sqrt{1-\rho^2}$. Substituting this series back into equation 23, and integrating each term separately, yields $\psi_{a^mb^n}=$ zero, if $m+n$ is odd. (This is for even powers of the signal all positive.) For m and n both even or both odd

$$\psi_{a^mb^n}=\sigma_a{}^m\sigma_b{}^n\Bigg[c_{m+n-1}\rho^m+$$

$$\frac{m(m-1)}{2!}\,c_{m+n-3}(1-\rho^2)\rho^{m-2}+$$

$$\frac{m(m-1)(m-2)(m-3)}{4!}\times$$

$$c_3c_{m+n-5}(1-\rho^2)^2\,\rho_{m-4}+\ldots+$$

$$\frac{m!}{(m-2q+2)!(2q-2)!}c_{2q-3}\times$$

$$c_{m+n-2q+1}(1-\rho^2)^{q-1}\rho^{m-2q+2}+\ldots+$$

$$\begin{cases} c_{m-1}\,c_{n-1}(1-\rho^2)^{\frac{m}{2}} \text{ if } m \text{ is even} \\ \text{or} \\ c_{m-2}\,c_n\rho(1-\rho^2)^{\frac{m-1}{2}} \text{ if } m \text{ is odd}\end{cases} \qquad (26)$$

Table I. CC as Function of ρ

m	n	$\sigma_a{}^{-m}\sigma_b{}^{-n}\psi_{a^mb^n}$
1	1	ρ*
3	1	3ρ
5	1	$3\times5\rho$
7	1	$3\times5\times7\rho$
m	1	$c_m\rho$
1	3	3ρ
3	3	$3(3\rho+2\rho^3)$
5	3	$3\times5(3\rho+4\rho^3)$
7	3	$3\times5\times7(3\rho+6\rho^3)$
m	3	$c_m(3\rho+(m-1)\rho^3)$
1	5	$c_5\rho$
3	5	$c_3(c_5\rho+20\rho^3)$
5	5	$c_5(c_5\rho+40\rho^3+4\times2\rho^5)$
7	5	$c_7(c_5\rho+60\rho^3+6\times4\rho^5)$
m	5	$c_m\left(c_5\rho+\frac{c_5(m-1)(5-1)}{2\times3}\rho^3+\frac{(m-1)(m-3)(5-1)(5-3)}{2\times4}\rho^5\right)$
1	7	$c_7\rho$
3	7	$c_3(c_7\rho+c_72\rho^3)$
5	7	$c_5\left(c_7\rho+c_7\frac{(5-1)(7-1)}{2\times3}\rho^3+c_7\frac{(5-1)(5-3)(7-1)(7-3)}{2\times4\times3\times5}\rho^5\right)$
7	7	$c_7\left(c_7\rho+c_7\frac{(7-1)(7-1)}{2\times3}\rho^3+c_7\frac{(7-1)(7-3)(7-1)(7-3)}{2\times4\times3\times5}\rho^5+\right.$ $\left.c_7\frac{(7-1)(7-3)(7-5)(7-1)(7-3)(7-5)}{7!}\rho^7\right)$
m, odd	n, odd	$c_mc_n\left(\rho+\frac{(m-1)(n-1)}{3!}\rho^3+\frac{(m-1)(m-3)(n-1)(n-3)}{5!}\rho^5+\ldots+\right)$ $\frac{\left(\frac{m-1}{2}\right)!\left(\frac{n-1}{2}\right)!2^{2k-2}\rho^{2k-1}}{\left(\frac{m+1}{2}-k\right)!\left(\frac{n+1}{2}-k\right)!(2k-1)!}$

* This term means $\rho_{ab}(\tau)=[\psi_{ab}(\tau)]/[\sigma_a\sigma_b]$.

There are k terms, where k equals $(m+1)/2$ or $(n+1)/2$, whichever is smaller.

Table II. CC as Function of τ

m	n	$\sigma_a{}^{-m}\sigma_n{}^{-n}\psi_{a^mb^n}$
2	2	$1+2\rho^2$*
4	2	$3(1+4\rho^2)$
6	2	$3\times5(1+6\rho^2)$
m	2	$c_{m-1}(1+m\rho^2)$
2	4	$c_3+12\rho^2$
4	4	$3(c_3+24\rho^2+8\rho^4)$
6	4	$3\times5(c_3+36\rho^2+48\rho^4)$
8	4	$c_7(c_3+48\rho^2+48\rho^4)$
m	4	$c_{m-1}(c_{n-1}+c_{n-1}\times2m\rho^2+m(m-2)\rho^4)$
2	6	$c_5+c_3\times6\rho^2$
4	6	$c_3(c_5+c_3\times12\rho^2+c_3\times8\rho^4)$
6	6	$c_5(c_5+c_3\times18\rho^2+c_3\times24\rho^4+6\times4\times2\rho^6)$
8	6	$c_7\left(c_5+c_3\times\frac{8\times6}{2}\rho^2+c_3\frac{8\times6\times6\times4}{4!}\rho^4+8\times6\times4\rho^6\right)$
m, even	n, even	$c_{m-1}c_{n-1}\left(1+\frac{mn}{2}\rho^2+\frac{m(m-2)n(n-2)}{4!}\rho^4+\frac{m(n-2)(m-4)n(n-2)(n-4)}{6!}\rho^6+\ldots+\right)$ $\frac{\left(\frac{m}{2}\right)!\left(\frac{n}{2}\right)!2^{2k}}{\left(\frac{m}{2}-k\right)!\left(\frac{n}{2}-k\right)!}\rho^{2k}$

* This term means $\rho_{ab}(\tau)=[\psi_{ab}(\tau)]/[\sigma_a\sigma_b]$.

There are k terms, where k equals m or n, whichever is smaller.

The CC from input to output of a single nonlinearity is also given by the column for n equals unity in Tables I and II. The input voltage is $\sigma_a=\sigma_b$, and $\rho_{ab}(\tau)$ is the input autocorrelation; m is the order of the power law nonlinearity. The autocorrelation of the output of this single nonlinearity is given by the row for $m=n=$ order of nonlinearity.

Either the specific tables, I and II, or a careful examination of the general series in the foregoing will reveal that

$$\frac{\psi_{a^mb^n}}{\sigma_a{}^m\sigma_b{}^n}=\frac{\psi_{a^nb^m}}{\sigma_a{}^n\sigma_b{}^m} \qquad (27)$$

The tables give the CC as a function of τ between the outputs of two nonlinear components. The a component is a power law of the mth order and the b component is a power law of the nth order. The inputs to each have gaussian probability distributions and are σ_a and σ_b volts rms respectively. The normalized CC between the inputs is $\rho_{ab}(\tau)$.

The power spectrum is the Fourier transform of the autocorrelation. The power in volts squared is $\psi(0)$. The power output of a nonlinearity is also given by these cross-power tables. For the self-power, set $a=b$, $m=n$, and $\rho=$unity. The increase in power is apparent. Furthermore, much of it is contributed by the higher powers of ρ, which have small time constants, and represent the higher harmonics generated by the nonlinearity. These higher harmonics are not correlated (in the usual sense) with the input, as the $n=1$ column shows. In fact, the only correlation with the input is a term proportional to the input autocorrelation.

Nonlinear Network

Fig. 3 shows the mechanism of action of a general nonlinearity expressed as a power series. The input θ_i is Gaussian with a power of σ^2 and a normalized autocorrelation of ρ.

The output θ_o is

$$\theta_o=n_1\theta_i+n_2\theta_i^2+n_3\theta_i^3+n_4\theta_i^4+\ldots \tag{28}$$

The cubic term delivers a component θ_3 to the output, where θ_3 is equal to $n_3\theta_i^3$. This component has an autocorrelation of

$$\psi_{33}=n_3^2\psi_{i^3i^3}=n_3^2(3\rho+2\rho^3)\sigma^6 \tag{29}$$

In a similar manner, the autocorrelation of each component in the output may be written from the previous table and multiplied by the gain coefficient squared.

All of the odd outputs are cross-correlated with each other. For example.

$$\psi_{35}=n_3n_5\psi_{i^3i^5}=n_3n_5\ 15(3\rho+4\rho^3)\sigma^8 \tag{30}$$

The even outputs are also cross-correlated with each other, but there is no correlation between the sums of the odds and the sums of the evens.

The CC between the input and the output of the entire system is due only to the odd terms. It is the sum of all of the $n-1$ terms in the table. Factoring out $\sigma\rho$ yields

$$\psi_{io}(\tau)=\sigma\rho(n_1\sigma+n_3c_3\sigma^3+n_5c_5\sigma^5+n_7c_7\sigma^7+\ldots)=N_1\sigma\rho(\tau)=\left(\frac{N_1}{\sigma}\right)\psi_{ii}(\tau) \tag{31A}$$

Equation 31(A) can serve as the definition of the equivalent gain N_1/σ.

This series may be written in terms of the moments of the distribution:

$$\psi_{io}=\rho\sigma^2\left[n_1+n_3\left(\frac{M_4}{\sigma^2}\right)+n_5\left(\frac{M_6}{\sigma^2}\right)+n_7\left(\frac{M_8}{\sigma^2}\right)+\ldots\right] \tag{31B}$$

Nonlinear Gain

The fundamental gain of a device is the ratio of the sinusoidal output divided by the sinusoidal input for a single frequency. When the signal is random, the fundamental gain is the cross-power spectral operator, which is the ratio of the cross power to the input power. For dynamic devices, the gain is a function of frequency. For nondynamic nonlinearities the gain is a function of the voltage amplitude.

The gain varies with the signal level and the coefficients of the higher powers. Let N_1 stand for the bracket in equation 31(A) which is a function of the n coefficients and σ. The Fourier transform of the cross-power spectral density of the nonlinearity is

$$\mathfrak{F}\psi_{io}(\tau)=\phi_{io}(\lambda)=\left(\frac{N_i}{\sigma}\right)\phi_{ii}(\lambda) \tag{32}$$

This includes all of the power terms in the output that are correlated with the input, and that have a bearing on the stability of a system when reintroduced into the input through linear feedback channels. The equivalent transference of a single nonlinearity in a linear system, with a Gaussian input, is N_1/σ. This is a variable, amplitude-dependent gain, without phase shift, and independent of frequency for a power series nonlinearity. The signs of the n coefficients determine whether the gain increases or decreases with amplitude. For stability studies this transference can be plotted as a scale along the 180-degree line in the L plane, (log gain versus phase), and marked with the rms level at the input to the nonlinearity.

The total output power of a nonlinearity can be derived from the autocorrelation or the self-power spectrum. There are CC between all of the even-power components and between all of the odd-power components. The autocorrelation of the output in Fig. 3 is the sum of all terms in the previous tables, each multiplied by the appropriate network coefficient. Notice that every CC term appears twice. Write all of these, and factor out the various powers of ρ. The coefficient of each power of ρ can be arranged in the order of increasing subscripts, and by induction it can be shown that every doubly infinite series is a square. The output autocorrelation is

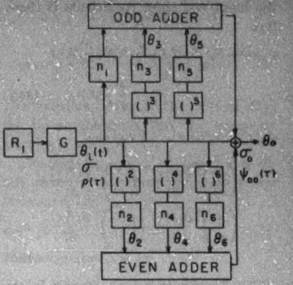


Fig. 3. General nonlinearity consisting of sum of all powers of input time function. There is no correlation between the output of odd adder and output of even adder

$$\begin{aligned}\psi_{oo}(\tau)={}&\rho(n_1\sigma+n_3c_3\sigma^3+n_5c_5\sigma^5+\ldots)^2+\\&\frac{\rho^3}{3!}(2n_3c_3\sigma^3+4n_5c_5\sigma^5+6n_7c_7\sigma^7+\ldots)^2+\\&\frac{\rho^5}{5!}(2\times4n_5c_5\sigma^5+4\times6n_7c_7\sigma^7+\ldots)^2+\\&\ldots+(n_2\sigma^2+n_4c_3\sigma^4+n_6c_5\sigma^6+\ldots)^2+\\&\frac{\rho^2}{2!}(2n_2\sigma^2+4n_4c_3\sigma^4+6n_6c_5\sigma^6+\ldots)^2+\\&\frac{\rho^4}{4!}(2\times4n_4c_3\sigma^4+4\times6n_6c_5\sigma^6+\\&6\times8n_8c_7\sigma^8+\ldots)^2+\ldots\end{aligned} \tag{33}$$

The coefficients of the even powers of ρ for zero average input will be defined as

$$N_k=\left[\frac{k}{2}!2^{\frac{k}{2}}n_kc_{k-1}\sigma^k+\frac{\left(\frac{k}{2}+1\right)!2^{\frac{k}{2}+1}}{1!2}n_{k+2}c_{k+1}\sigma^{k+2}+\frac{\left(\frac{k}{2}+2\right)!2^{\frac{k}{2}+2}}{2!2^2}\times n_{k+4}c_{k+3}\sigma^{k+4}+\ldots\right] \tag{34}$$

The coefficients of the odd powers of ρ for zero average input will be defined as

$$N_r=\left[\left(\frac{r-1}{2}\right)!2^{\frac{r-1}{2}}n_rc\sigma_r{}^r+\frac{\left(\frac{r+1}{2}\right)!2^{\frac{r+1}{2}}}{1!2^2}n_{r+2}c_{r+2}\sigma^{r+2}+\frac{\left(\frac{r+3}{2}\right)!2^{\frac{r+3}{2}}}{2!2^2}n_{r+4}c_{r+4}\sigma^{r+4}+\ldots\right] \tag{35}$$

The output autocorrelation in terms of

the input and these coefficients is therefore

$$\psi_{oo}(\tau)=N_0^2+\rho N_1^2+\frac{\rho^2}{2!}N_2^2+\frac{\rho^3}{3!}N_3^2+\frac{\rho^4}{4!}N_4^2+\ldots \quad (36A)$$

$$\psi_{oo}(\tau)=+N_0^2+\left(\frac{N_1}{\sigma}\right)^2\psi_{ii}(\tau)+\left(\frac{N_2}{\sigma^2}\right)^2\frac{\psi_{ii}^2(\tau)}{2!}+\left(\frac{N_3}{\sigma^3}\right)^2\frac{\psi_{ii}^3(\tau)}{3!}+\left(\frac{N_4}{\sigma_4}\right)^2\frac{\psi_{ii}^4(\tau)}{4!}+\left(\frac{N_5}{\sigma^5}\right)^2\frac{\psi_{ii}^5(\tau)}{5!}+\ldots \quad (36B)$$

The Fourier transform of this is the power spectrum. The output power in volts squared is expressed in equation 36(A) when τ is zero and ρ is unity:

$$\psi_{oo}(0)=N_0^2+N_1^2+\frac{N_2^2}{2!}+\frac{N_3^2}{3!}+\frac{N_4^2}{4!}+\ldots \quad (37)$$

This may contain considerable harmonic power, since N_1^2 is the only term resulting from the useful output power, which correlates with the input.

The N^2 coefficients are powers in volts squared of combinations of harmonics. The significance of the N coefficients may be shown by considering, for example, N_3^2. This determines the magnitude of ψ_{ii}^3 in the output. To find these specific terms, expand the cube of the autocorrelation.

$$\psi_{ii}^3=\left(a\epsilon^{-\frac{|\tau|}{T_1}}+b\epsilon^{\frac{|\tau|}{T_2}}+c\epsilon^{-\frac{|\tau|}{T_3}}+\ldots\right)^3$$
$$=a^3\epsilon^{-\frac{3|\tau|}{T_1}}+b^3\epsilon^{-\frac{3|\tau|}{T_2}}+3a^2b\epsilon^{-|\tau|\left(\frac{2}{T_1}+\frac{1}{T_2}\right)}+3ab^2\epsilon^{-|\tau|\left(\frac{1}{T_1}+\frac{2}{T_2}\right)}+\ldots \quad (38)$$

N_3 therefore controls the magnitude of the third harmonics and the third-order beat frequencies, i.e., one frequency plus twice another frequency.

In the power density spectrum, N_3 controls the magnitude or residues of an additional set of poles which are added to the original ones. Additional poles are found at three times the radial distance from the origin to each of the original poles. Additional poles are at all possible combinations of the sum of one original position plus twice the radial distance from the origin to each of the other poles. In a similar manner, N_5 controls the magnitude of an additional set of fifth-order beat frequency poles in the output spectrum.

Equation 36 can be used to calculate the output power of a nonlinear element. The first term is the direct current from rectification. The second term is all of the output power which correlates with the input, and the other terms are the harmonic powers. In a feedback system, if the nonlinearity is near the system output, these harmonic terms contribute to the error. Since they are introduced within the loop, at a high power-level point, they will be degenerated by the loop gain. They may be considered as noise inputs, whose amplitudes correlate with the signal amplitude at the input to the nonlinearity. The harmonic power does not have a Gaussian distribution. It has components which are the cubes and higher powers of a Gaussian distribution. If these terms are more objectionable than an equal amount of power in an error with Gaussian distribution, then the minimum error power is not the best criterion of design to use.

Feedback Loop with Nonlinear Amplifier

The loop gain equations are derived by considering a signal going through the loop, arriving back at the input, and subtracting from the input. In Fig. 4, the nonlinearity in the forward branch is represented by its cross-power operator G and the harmonic distortion signal θ_h with power ϕ_{hh} (λ). At the subtractor, $\theta_i-F\theta_0$ equals θ_e, and cross powers to the input are

$$\phi_{ie}=\frac{\phi_{io}}{G}=\phi_{ii}-F_{io} \quad (39)$$

since θ_h does not correlate with θ_i. Solving for ϕ_{io} yields

$$\phi_{io}=\frac{G}{1+FG}\phi_{ii} \quad (40)$$

The equivalent closed-loop transference is a function only of the cross-power operators G and F. For a nonlinearity, the cross-power operator is given by equation 31 as (N_1/σ). This could be thought of as a "linearization" for Gaussian signals, or as the "gain" of a nonlinearity.

If ϕ_{hh} is the harmonic power appearing at the output when the loop is open, and if it is tested at the normal operating power levels and probability distributions, then closing the loop will cause the harmonic component of the output cross-power to be

$$\phi_{ho}=\frac{1}{1+FG}\phi_{hh} \quad (41)$$

The output self-power due to the input is

$$\frac{G}{1+FG}\overline{\frac{G}{1+FG}}\phi_{ii}(\lambda) \quad (42A)$$

The bar above the operator means the complex conjugate as a function of the variable $\lambda=\omega+ja$.

The output self-power due to the harmonic generation is

$$\frac{1}{1+FG}\overline{\frac{1}{1+FG}}\phi_{hh}(\lambda) \quad (42B)$$

The total output power is the sum of equations 42(A) and (B).

$$\phi_{oo}=\frac{1}{(1+FG)\overline{(1+FG)}}(\bar{G}\bar{G}\phi_{ii}+\phi_{hh}) \quad (42C)$$

The transferences must be evaluated for the case when the nonlinearity receives a probability distribution which is the same as that produced by the signal plus the harmonic distortion, as it appears going through the loop.

Error Power Due to a Nonlinearity

An amplifier which is supposed to have a gain of n_1 may have a saturation curve which is describable by a power series. The error voltage θ_e at the output will be the input times a linear constant gain of n_1, minus the output of a power-series nonlinearity. The error autocorrelation is

$$\psi_{ec}=n_1^2\psi_{ii}-n_1\psi_{io}-n_1\psi_{oi}+\psi_{oo} \quad (43)$$

Collect the terms with the same spectrum as the input.

$$\psi_{ee}=n_1^2\sigma^2\rho-2n_1\sigma N_1\rho+N_1^2\rho+\left[N_0^2+\frac{\rho^2}{2!}N_2^2+\frac{\rho^3}{3!}N_3^2+\ldots\right] \quad (44A)$$

$$\psi_{ee}=\rho(N_1-n_1\sigma)^2+\left[N_0^2+\frac{\rho^2}{2!}N_2^2+\frac{\rho^3}{3!}N_3^2+.\right] \quad (44B)$$

The error power in volts squared follows directly by substituting τ equals zero, ρ equals one.

$$\psi_{ee}(0)=(n_3c_3\sigma^3+n_5c_5\sigma^5+\ldots)^2+\left[N_0^2+\frac{N_2^2}{2!}+\frac{N_3^2}{3!}+\ldots\right] \quad (45)$$

The second bracket is all harmonic power; the first is error power due to the change in gain from n_1 to N_1/σ.

In case the nonlinearity has only cubic distortion, the error power is $5(n_3c_3\sigma^3)^{2/3}$ volts squared. In case it has only square

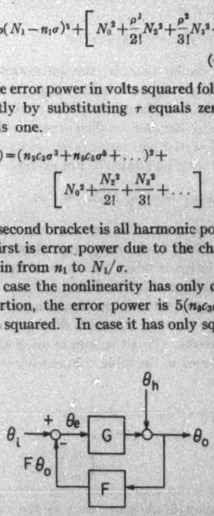


Fig. 4. Nonlinearity in forward branch of feedback system represented as linear transference G and harmonic power input θ_h

Vol.
Pt. 1
1959

distortion, the error power is $3(n_2\sigma^2)^2$ volts squared, part of which is direct current due to rectification.

In equation 32, the equivalent nonlinear gain is given as N_1/σ. If the error power at the output of a nonlinearity is computed with respect to a linear branch with gain of N_1/σ, then the coefficient of the first power of ρ in the foregoing is zero, and the error power is

$$\psi_{ee}(0)=\left[N_0^2+\frac{N_2^2}{2!}+\frac{N_3^2}{3!}+\ldots\right] \tag{46}$$

If one minimizes the error power between the outputs of a nonlinearity and a linear variable gain device, he arrives at the foregoing error power, corresponding to a linear gain setting of N_1/σ.

Filtering Effect of Nonlinearities and Linearities

A Gaussian signal going through a linear filter is still Gaussian when it emerges from the output. If a non-Gaussian signal is passed through several filters, so that the output power is markedly reduced owing to the reduction in bandwidth, the resulting signal is much nearer gaussian, and as the bandwidth approaches zero, the distribution approaches a Gaussian one. In general, then, if a nonlinearity occurs in or after the high gain in a control system, the subsequent filtering of the harmonics (due to the closed loop characteristics), before they reach the output, will change their distribution to an almost Gaussian one. The power spectrum of the harmonics at any point in the system beyond the nonlinearity can be determined by starting with the Fourier transform of equation 36, and operating on it with the linear transferences of the successive linear members. Because of the filtering action, it is usually permissible to determine nonlinear transferences for the assumption of a Gaussian input.

If the nonlinearity definitely receives a non-Gaussian signal, there are two approaches to the problem. One is to measure the crosspower on the actual equipment as a function of different input levels. The other is to break up the input signal into an equivalent Gaussian signal and a nonlinear operator; the latter approach will be discussed in a subsequent section.

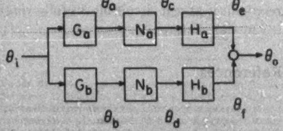


Fig. 5. Sum of two signals from parallel branches with nonlinearities N_a and N_b imbedded between linear filters

A nonlinearity is an amplitude filter. As such, it changes both the distribution and the spectrum. If the filter selects a narrow band of amplitudes, the output looks like a random sequence of step functions. The ultimate in narrow-band amplitude filtering is a probability density distribution with delta functions at only two amplitude values. This approaches the single time constant spectrum of clipped thermal noise. The time constant is zero if the thermal noise has infinite bandwidth. In other words, a flat spectrum into a narrow-band amplitude filter emerges as a flat spectrum. In general, because of the harmonic generation of a nonlinearity, a narrow spectrum input will have considerable energy removed from the spectrum and reintroduced at higher frequencies. This usually makes the output spectrum much flatter than the input.

As a general rule, linear filters tend to produce a Gaussian amplitude distribution, and a "peaked" power spectrum. Nonlinearities tend to produce a "flat" power spectrum and peaked amplitude distribution. If the nonlinear peaking of a Gaussian amplitude distribution occurs along the skirts of the curve, it may at first sight appear like a flattening of the center.

Multiple Nonlinearities

If two nonlinearities are placed in parallel, the corresponding n's of each should be added for use in equations 31 through 37. Since the N's are additive combinations of the n's, they too may be added. If in Fig. 5 all the G's and H's are unity

$$\theta_o=\theta_{ao}+\theta_{bo} \tag{47}$$

$$\theta_{ao}=n_{a1}\theta_i+n_{a2}\theta_i^2+n_{a3}\theta_i^3+\ldots$$

$$\theta_{bo}=n_{b1}\theta_i+n_{b1}\theta_i^2+n_{b3}\theta_i^3+\ldots \tag{48}$$

$$N_1=N_{a1}+N_{b1}$$

$$N_2=N_{a2}+N_{b2},\ \text{etc!} \tag{49}$$

Fig. 5 shows two nonlinearities in parallel branches, but isolated from each other by linear dynamic elements G and H.

The output power and cross power of this circuit can be written, with the caution that higher harmonics of the nonlinearities are correlated with each other.

The cross power is the sum of the cross powers of each branch:

$$\phi_{io}(\lambda)=H_a\left(\frac{N_{a1}}{\sigma_a}\right)G_a\phi_{ii}+H_b\left(\frac{N_{1b}}{\sigma_b}\right)$$
$$G_b\phi_{ii}=\left(\frac{H_aG_aN_{a1}}{\sigma_a}+\frac{H_bG_bN_{b1}}{\sigma_b}\right)\phi_{ii}(\lambda) \tag{50}$$

The transference has all the poles of all of the linear elements, but the zeros are amplitude-sensitive.

The first step toward obtaining the output power is to write the cross power between θ_a and θ_b.

$$\phi_{ab}=\bar{G}_aG_b\phi_{ii} \tag{51}$$

The normalized CC is

$$\rho_{ab}(\tau)=\frac{\psi_{ab}(\tau)}{\sigma_a\sigma_b}=\frac{\mathfrak{F}^{-1}\bar{G}_aG_b\phi_{ii}(\lambda)}{\sigma_a\sigma_b} \tag{52}$$

The CC between θ_c and θ_d is given by Tables I and II. Each term in these tables is the correlation between the mth power of θ_a and the nth power of θ_b. Each term should therefore be multiplied by the appropriate coefficients in N_a and N_b, namely, $n_{am}n_{bn}$.

The CC ψ_{cd} is the sum of all terms in these tables in the form of $n_{am}n_{bn}\psi_{ambn}$. Each term contains the product $\sigma_a{}^m\sigma_b{}^n$. Therefore, in the sum, one need only attach the coefficient n_{am} to each $\sigma_a{}^m$, and the coefficient n_{bn} to each $\sigma_b{}^n$. All terms in the table are added together, and coefficients of the powers of ρ collected together. When this is done and the coefficients are identified, they have the following simple form:

$$\psi_{cd}=N_{ao}N_{bo}+\rho_{ab}N_{a1}N_{b1}+\rho_{ab}{}^2\frac{N_{a2}N_{b2}}{2!}+$$
$$\rho_{ab}{}^3\frac{N_{a3}N_{b3}}{3!}+\rho_{ab}{}^4\frac{N_{a4}N_{b4}}{4!}+\ldots \tag{53}$$

The N's are given by equations 34 and 35, for zero average input. They are the same N's that are used for the autocorrelation at points c and d.

$$\psi_{cc}=N_{ao}{}^2+\rho_{aa}N_{a1}{}^2+\rho_{aa}{}^2\frac{N_{a2}{}^2}{2!}+$$
$$\rho_{aa}{}^3\frac{N_{a3}{}^2}{3!}+\ldots \tag{54}$$

The output power spectrum of the sum of the two parallel branches in Fig. 5 is

$$\phi_{oo}=H_aH_a\mathfrak{F}\psi_{cc}+H_bH_b\mathfrak{F}\psi_{dd}+$$
$$H_aH_b\mathfrak{F}\psi_{cd}+H_bH_a\mathfrak{F}\psi_{dc} \tag{55}$$

When two nonlinearities are placed in series, the resultant n's are given by the substitution of the first signal polynomial into the second. The order of power flow is quite important.

In Fig. 6(A)

$$\theta_b=n_{a1}\theta_i+n_{a2}\theta_i^2+n_{a3}\theta_i^3+\ldots \tag{56}$$

$$\theta_o=n_{b1}\theta_b+n_{b2}\theta_b^2+n_{b3}\theta_b^3+\ldots \tag{57}$$

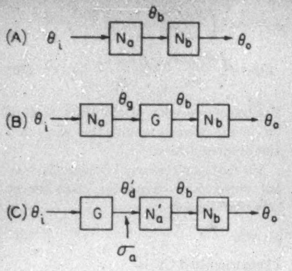


Fig. 6. Systems with nonlinearities in series. A—One nonlinearity immediately following another without filtering: can be combined directly. B—Linear dynamic filter following one nonlinearity and preceding another. C—Transformation of (B) by moving linear filter to input so that two resultant nonlinearities can be combined

Substituting and collecting terms yields

$$\theta_o=\theta_i(n_{b1}n_{a1})+\theta_i^2(n_{b1}n_{a2}+n_{b2}n_{a1}^2)+ \theta_i^3(n_{b1}n_{a3}+n_{b2}(2n_{a1}n_{a2})+n_{b3}n_{a1}^3)+ \theta_i^4[n_{b1}n_{a4}+n_{b2}(n_{a2}^2+2n_{a1}n_{a3})+ n_{b3}(3n_{a1}^2n_{a2})+n_{b4}n_{a1}^4]+\theta_i^5[n_{b1}n_{a5}+ n_{b2}(2n_{a1}n_{a4}+2n_{a2}n_{a3})+n_{b3}(3n_{a1}n_{a2}^2+ 3n_{a1}^2n_{a3})+n_{b4}(4n_{a1}^3n_{a2})+n_{b5}n_{a1}^5]+ \theta_i^6[n_{b1}n_{a6}+n_{b2}(n_{a3}^2+2n_{a1}n_{a5}+ 2n_{a2}n_{a4})+n_{b3}(n_{a2}^3+3n_{a1}^2n_{a4})+ n_{b4}(6n_{a1}^2n_{a2}^2+4n_{a1}^3n_{a3})+ n_{b5}(5n_{a1}^4n_{a2})+n_{b6}n_{a1}^6]+\ldots \quad (58)$$

Therefore n_{c1} is $n_{b1}n_{a1}$; n_{c2} is $(n_{b1}n_{a1}^2+n_{b1}n_{a2})$; etc. Using the N_c coefficients eliminates the necessity of computing the rms value σ_b.

When two nonlinearities are separated by a linear element as in Fig. 6(B), the cross power is not much more difficult, but the self-power is. For the case of a d-c level change at θ_g (so that both nonlinearities receive zero average input) and the assumption of sufficient filtering (so that the inputs to both nonlinearities are Gaussian),

$$\phi_{io}=\left(\frac{N_{b1}}{\sigma_b}\right)G\left(\frac{N_{a1}}{\sigma_a}\right)\phi_{ii} \quad (59)$$

$$\phi_{gg}=\mathfrak{F}\left[\left(\frac{N_{a1}}{\sigma_a}\right)^2\psi_{ii}+\left(\frac{N_{a2}}{\sigma_a^2}\right)^2\frac{\psi_{ii}^2}{2!}+ \left(\frac{N_{a3}}{\sigma_a^3}\right)\frac{\psi_{ii}^3}{3!}+\ldots \quad (60)$$

$$\psi_{oo}=\left(\frac{N_{b1}}{\sigma_b}\right)^2\mathfrak{F}^{-1}G\bar{G}\phi_{gg})+\left(\frac{N_b^2}{\sigma_b^2}\right)^2\times \frac{(\mathfrak{F}^{-1}G\bar{G}\phi_{gg})^2}{2!}+\left(\frac{N_{b3}}{\sigma_b^3}\right)^2\frac{(\mathfrak{F}^{-1}G\bar{G}\phi_{gg})^3}{3!}+ \ldots \quad (61)$$

If G provides so little filtering that the input to the second nonlinearity cannot be considered Gaussian, then the circuit in Fig. 6(B) should be converted to the arrangement of Fig. 6(C). To derive the N_a' operator, consider that there are correction factors on the n_a coefficients to account for the attenuation due to G. First, put G ahead of N_a. Next, solve for ϕ_{aa}, ψ_{aa}, σ_a, and ρ_a. Write separately the power output at θ_b for each term in the nonlinearity N_a', and set each equal to the old power output for the term of the same order. For example, the cube has

$$\psi_b'(\tau)=(n_{a3}'c_3\sigma_a^3)^2\rho_a(\tau)+ (2n_{a3}'c_{3a}\sigma^3)^2\frac{\rho_a^3(\tau)}{3!} \quad (62)$$

The previous value was

$$\psi_b(\tau)=(n_{a3}c_3\sigma_i^3)^2\rho_0(\tau)+ (2n_{a3}c_3\sigma_i^3)^2\frac{\rho_i^3(\tau)}{3!} \quad (63)$$

Setting these equal for τ of zero yields

$$n_{a3}'\sigma_a^3=n_{a3}\sigma_i^3 \quad (64)$$

In a similar manner, every n coefficient will be increased by the corresponding power of the ratio of the old to the new rms input of N_a. But note that since every n coefficient appears only times the corresponding power of σ, the values of N_{a1}', N_{a2}', N_{a3}', etc., are identical to the old values of N_{a1}, N_{a2}, N_{a3}, etc., based on σ_i.

In a similar manner, it can be shown that using the identical nonlinear transference as before will yield the correct cross powers (watts, not spectrum) and correct total output power. It does not give the correct power spectrum or the autocorrelation and, consequently, the output of the following filter or nonlinearity will not be correct. The rule for moving a linear filter from a position following a nonlinearity to a position head of it, is to change their positions without any change in the linear transference, and with the N's of the nonlinear transference based on its old rms input, (σ_i), operating on its new spectral input (ψ_{aa}'/σ_a). From this new form, the new nonlinear coefficients can be combined with those of N_b in Fig. 6(C), with the use of equation 58, to obtain an over-all single nonlinear transference.

Non-Gaussian Distributions

The response of smooth nonlinearities to Gaussian signals has been analyzed. The equivalent linear gain, the cross-power spectrum operator, and the output self-power spectrum can all be obtained easily. An approximation to the response to non-Gaussian distributions can be obtained by describing the input signal in terms of the mechanism by which it could be made. In Fig. 6(B), the signal θ_b is non-Gaussian owing to the harmonic generation of N_a. This signal could be described accurately in terms of the mechanism of N_a followed by G. However, for convenience in using the description to calculate actual responses, the circuit of Fig. 6(C) is preferable. Placing G first, followed by the nonlinearity N_a' as a signal description, permits the immediate combination of the nonlinear terms in the signal with the nonlinear terms in the circuit. Exactly this is done in linear analysis when a signal is described by its Laplace transform, so that it can be combined with the Laplace transform of the linear circuit, to obtain the transform of the output. The signal and the circuit are both described in the same language, by a pattern of poles and zeros in the s plane.

Here, both the signal and the circuit must be described in the same language, in this case by N operators. The signal description then takes the form of equation 36, where its measured autocorrelation function is expanded in a power series of an autocorrelation function that has no beat frequencies. Notice that the terms of equation 38 can be identified as the beat frequencies and harmonics of a fundamental correlation. When harmonic terms like these are apparent in a measured autocorrelation, a series like equation 36 should be chosen, and by setting corresponding terms equal, the values of Ψ_{ii}/σ_i and N_1, N_2, N_3, etc., may be determined. From the N's, the n's may be obtained algebraically. The signal may now be considered an original Gaussian distribution, tailored first by a linear filter G, and second by a power series N_a'; see Fig. 6(C). The block representing the nonlinear part of the signal may be combined with the circuit nonlinearity with the use of equation 58, and the output of the combination obtained directly from the known Gaussian input.

Whenever possible, a signal description should be in the same kind of mathematics as the circuit description. For a random signal, the best method is to represent it as real circuit blocks following a Gaussian, flat, random source.

References

1. NONLINEAR CONTROL SYSTEMS WITH STATISTICAL INPUTS, **Richard C. Booton, Jr.** *Report no. 61,* Dynamic Analysis and Control Laboratory, Massachusetts Institute of Technology, Lexington, Mass.

2. NONLINEAR SERVOMECHANISMS WITH RANDOM INPUTS, **Richard C. Booton, Jr.** *Report no. 70,*

Dynamic Analysis and Control Laboratory, Massachusetts Institute of Technology, Aug. 30, 1953.

3. Trends in Feedback Systems, **J. M. Otto Smith.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., CT-1 no. 1, Mar. 1954, pp. 2–8.

4. Nonlinear Control Systems with Random Inputs, **R. C. Booton, Jr.** *Ibid.*, Institute of Radio Engineers, CT-1 no. 1, Mar. 1954, pp. 9–18.

5. A Compensating Network Theory for Feedback Control Systems Subject to Saturation, **G. C. Newton, Jr.** *Report no. 22*, Servomechanisms Laboratory, Massachusetts Institute of Technology, June 5, 1950.

6. The Effects of Noise on Rate-Limited Systems, **L. A. Payne, R. G. Bowden, I. H. Van Horn, Jr.** *Report no. GER-2328*, Goodyear Aircraft Corporation, Akron, Ohio, Oct. 2, 1950.

7. A Digital Electronic Correlator, **H. E. Singleton.** *Technical Report no. 152*, Research Laboratory of Electronics, Massachusetts Institute of Technology, Feb. 21, 1950.

8. A Short-Time Correlator for Speech Waves, **P. E. A. Cowley, R. M. Fano, B. L. Basore.** *Technical Report no. 174*, Research Laboratory of Electronics, Massachusetts Institute of Technology, Feb. 26, 1951.

9. A Nonlinear Statistical Filter, **A. W. Sullivan, J. M. Barney.** Engineering and Industrial Experiment Station, University of Florida, Gainesville, Fla., Oct. 1951.

10. Experimental Study of Nonlinear Devices by Correlation Methods, **L. Weinberg, L. G. Draft.** *Technical Report no. 178*, Research Laboratory of Electronics, Massachusetts Institute of Technology, Jan. 20, 1951.

11. Statistical Errors in Measurements on Random Time Functions, **W. B. Davenport, Jr., R. A. Johnson, D. Middleton.** *Journal of Applied Physics*, New York, N. Y., vol. 23, no. 4, Apr. 1952, pp. 377–88.

12. Correlator for Low Frequencies, **V. J. Guethlen, R. W. Hix.** Goodyear Aircraft Corporation, Akron, Ohio, Nov. 1952.

13. Methods of Obtaining Amplitude-Frequency Spectra, **A. E. Hastings.** *Review of Scientific Instruments*, New York, N. Y., vol. 23, no. 7, July, 1952, p. 344.

14. Device for Computing Correlation Functions, **A. E. Hastings, J. E. Meade.** *Ibid.*, p. 347.

15. An Analog Electronic Correlator, **J. F. Reintjes.** *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 7, 1952, pp. 390–400.

16. A Simple Computer for Automatically Plotting Correlation Functions, **Allan H. Schooley.** *NRL Reprint no. 90-52*, Naval Research Laboratory, Washington, D. C., Dec., 1952.

17. A Five-Channel Electronic Analog Correlator, **M. J. Levin, A. F. Reintjens.** *Proceedings*, National Electronics Conference, vol. 8, Jan., 1953, pp. 647–59.

Spectral Output of Piecewise Linear Nonlinearity

OTTO J. M. SMITH
MEMBER AIEE

THE TRANSFERENCE or effective gain of a piecewise linear component is derived. The output autocorrelation function of a clipper, a dead zone, and a general piecewise linear nondynamic component is obtained by a double integration of the second probability density. The output power is a sum of Gaussian error integrals. This can be divided into fundamental or correlated power, and distortion power due to lack of desired signal or presence of harmonics.

Saturation or Clipping

A large class of nonlinearities includes those with sharp discontinuities in the output-input relationship. In this class are dead zone, saturation, and the approximations to any nonlinear curve by a sequence of straight-line segments. The last-mentioned may be thought of as the sum of several dead zones and saturations. The cross-power transferences and output power operators will be derived for these nonlinearities by a double integration of the second probability density. This method has also been applied to power series nonlinearities.[1–6] An output transducer can only deliver a finite maximum power or torque. The error power due to this "clipping" or saturation should be balanced against the cost of the transducer to obtain the best compromise. The output error as a function of the clipping level for random signals will be derived.

The input-output relation of a saturating element is shown in Fig. 1(A), where θ is the output variable, and x is the input variable. The proportional region has unity slope; k is the clipping level in terms of the input. Mathematically, the output is

$$\begin{aligned}\theta(x)&=k,\ k<x\\ \theta(x)&=x,\ -k<x<k\\ \theta(x)&=-k,\ x<-k\end{aligned}\tag{1}$$

The cross correlation (CC) between input and output is the double integral over all values of the input and output of the product of the input, times the output, times the second (joint) probability density.

$$\psi_{io}(\tau)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}w\theta(x)\left(\frac{\partial p}{\partial w}\,\frac{\partial p}{\partial x}\right)dwdx\tag{2}$$

The second probability density for a Gaussian input is given by

$$\left(\frac{\partial p}{\partial w}\,\frac{\partial p}{\partial x}\right)=\frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}}\times \exp\left(-\frac{w^2-2\rho wx+x^2}{2(1-\rho^2)\sigma^2}\right)\tag{3}$$

where σ is the rms value of the input and $\rho(\tau)$ is the normalized autocorrelation equal to $\psi_{ii}(\tau)/\sigma^2$. With the following change of variable, the exponent will appear as a complete square, and the integral with respect to u can be obtained immediately:

$$z=\frac{x}{\sigma}\tag{4}$$

$$u=\frac{\frac{w}{\sigma}-\rho\frac{x}{\sigma}}{\sqrt{2}\sqrt{1-\rho^2}}\tag{5}$$

$$\psi_{io}(\tau)=\frac{\sigma^2\sqrt{1-\rho^2}}{\pi}\int_{-\infty}^{+\infty}\frac{\theta(z)}{\sigma}\epsilon^{-\frac{z^2}{2}}\times dz\int_{-\infty}^{\infty}\left(u+\frac{\rho z}{\sqrt{2}\sqrt{1-\rho^2}}\epsilon^{-u^2}\right)du\tag{6}$$

Performing the last integration yields

$$\psi_{io}(\tau)=\frac{\sigma^2\rho(\tau)}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\frac{\theta(z)}{\sigma}\times z\epsilon^{-\frac{z^2}{2}}dz\tag{7}$$

In equation 7, notice that all functions of time have come outside of the integral.

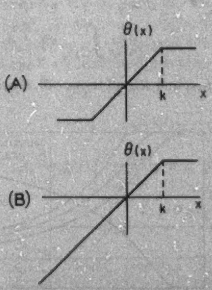


Fig. 1. Input-output relation of clipper or saturating component

A—Symmetrical
B—Single-sided clipping

Paper 59-264, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted December 30, 1957; made available for printing December 17, 1958.

Otto J. M. Smith is with the University of California, Berkeley, Calif.

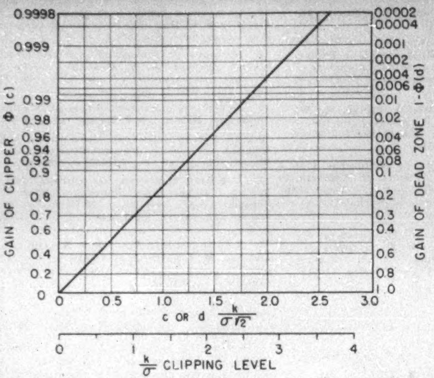


Fig. 2. Statistical gain of symmetrical clipper and of symmetrical dead zone as function of ratio of clipping level to rms signal

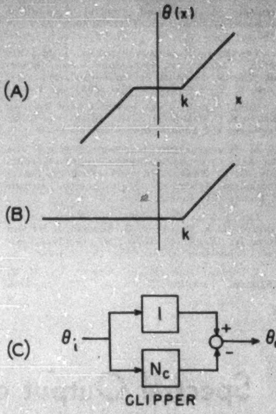


Fig. 4. Dead zone. A—Input-output relation for symmetrical dead zone. B—One-sided dead zone. C—Exact equivalent circuit for dead zone

This can therefore be considered a complete general equation for the calculation of the CC of any nondynamic function with Gaussian input. The Fourier transform can be taken before integrating. The transference then becomes only an amplitude-dependent gain term, for any kind of nondynamic nonlinearity. The cross-power spectrum is

$$\phi_{io}(\omega)=\phi_{ii}(\omega)\times\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\frac{\theta(z)}{\sigma}z\epsilon^{-\frac{z^2}{2}}\,dz \tag{8}$$

where z is the normalized input signal.

For the case of symmetrical clipping, $\theta(z)/\sigma$ equals z between values of z of $\pm k/\sigma$, and equals k/σ outside of this range. Therefore, when these values are substituted, the transference is

$$\frac{\phi_{io}}{\phi_{ii}}=\frac{k}{\sigma\sqrt{2\pi}}\int_{-\infty}^{-k/\sigma}\epsilon^{-\frac{z^2}{2}}z\,dz+\frac{k}{\sigma\sqrt{2\pi}}\int_{+k/\sigma}^{+\infty}\epsilon^{-\frac{z^2}{2}}z\,dz+\frac{1}{\sqrt{2\pi}}\int_{-k/\sigma}^{+k/\sigma}z^2\epsilon^{-\frac{z^2}{2}}\,dz \tag{9}$$

The first and second terms cancel because the first term is a negative area, and the second an equal positive area.

Integrating the last term by parts yields

$$\frac{\phi_{io}(\omega)}{\phi_{ii}(\omega)}=\frac{1}{\sqrt{2\pi}}\left[-z\epsilon^{-\frac{z^2}{2}}\right]_{-k/\sigma}^{+k/\sigma}+\frac{1}{\sqrt{2\pi}}\int_{-k/\sigma}^{+k/\sigma}\epsilon^{-\frac{z^2}{2}}\,dz=\Phi\left(\frac{k}{\sigma\sqrt{2}}\right) \tag{10}$$

$\Phi(x)$ is defined as the Gaussian error integral or probability integral:

$$\Phi(x)=\frac{2}{\sqrt{\pi}}\int_0^x\epsilon^{-u^2}\,du \tag{11}$$

This is an important fundamental result. It states that the transference of a symmetrical clipper is a cross-power operator equal to the probability integral of a kind of normalized clipping level. The peak value of the clipping level is k.

Consider $\sqrt{2}\sigma$ as an "equivalent" peak input signal. Then the ratio of the clipping voltage to the equivalent peak input, or the ratio of the "effective" clipping voltage $k/\sqrt{2}$ to the effective or rms input voltage, is defined as the normalized clipping level c.

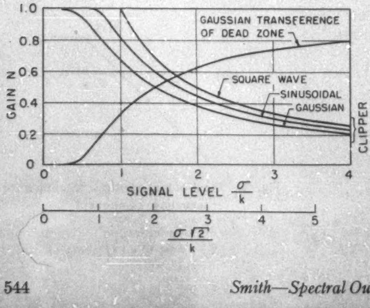


Fig. 3. Gain of clipper as function of type of input signal. For random Gaussian input, statistical gain is ratio of cross-power spectrum to input power spectrum. For analytical functions, fundamental harmonic component gain is shown

$$c=\frac{k}{\sigma\sqrt{2}} \tag{12}$$

The spectral transference is the Gaussian error integral of c. Since the clipping transference is not a time function, it is equally applicable to CCs as well as to cross powers.

$$\psi_{io}(\tau)=\Phi(c)\psi_{ii}(\tau) \tag{13}$$

This transference is plotted in Figs. 2 and 3 as a function of clipping level and signal level respectively. It can be considered a nondynamic amplitude-dependent gain term for the calculation of loop gain for stability studies. The component of the output power which correlates with the input is

$$\Phi^2(c)\psi_{ii}(0)=\sigma^2\Phi^2(c)\text{ watts} \tag{14}$$

There are, in addition, a number of harmonic power components, which will be derived later.

Fig. 1(B) shows an unsymmetrical clipper. To obtain its transference, change the lower limit of the last integral in equation 9 to $-\infty$, and omit the first integral. The unsymmetrical transference then becomes

$$\frac{\phi_{io}}{\phi_{ii}}=\frac{1}{2}+\frac{1}{2}\Phi(c) \tag{15}$$

Dead Zone

A dead zone component is exactly the same as one minus a clipping component; see Fig. 4. The cross-power transference and CC for a symmetrical dead zone are therefore

Vol.
Pt. 1
1959

$$\frac{\phi_{io}(\omega)}{\phi_{ii}(\omega)}=\frac{\psi_{io}(\tau)}{\psi_{ii}(\tau)}=1-\Phi\left(\frac{k}{\sigma\sqrt{2}}\right)=1-\Phi(d) \qquad (16)$$

A normalized dead zone d is defined as 0.707 times the peak dead zone in one direction divided by the rms input. This is true of a symmetrical dead zone with a Gaussian input. The dead-zone gain is also shown in Figs. 2 and 3. Notice that the additive relationship between the dead-zone transferences and the clipping transference, which is apparent for periodic inputs, also holds for random inputs.

An unsymmetrical dead zone is shown in Fig. 4(B). Its transference is exactly one minus that for an unsymmetrical clipper:

$$\frac{\phi_{io}(\omega)}{\phi_{ii}(\omega)}=\frac{1}{2}-\frac{1}{2}\Phi(d) \qquad (17)$$

General Discontinuous-Slope Nonlinearity

A large class of nonlinearities can be represented by straight-line segments for the input-output relationship. The transference of complicated devices can be written down immediately with the tools just derived. The symmetrical curve in Fig. 5(A) is the sum of a linear function and several dead zones.

At each point where the slope changes, the new dead-zone term contributes the change in slope. For abbreviation, slopes will be designated by m, the tangent of the angle, and probability integrals by

$$\Phi_3\equiv\Phi(d_3) \qquad (18)$$

The gain is

$$\frac{\phi_{io}}{\phi_{ii}}=m_1+(m_2-m_1)(1-\Phi_1)+ (m_3-m_2)(1-\Phi_2)+(m_4-m_3)(1-\Phi_3)+\ldots \qquad (19)$$

$$\frac{\phi_{io}}{\phi_{ii}}=m_1\Phi_1+m_2(\Phi_2-\Phi_1)+m_3(\Phi_3-\Phi_2)+\ldots \qquad (20)$$

The gain is the sum of each slope times the difference of the normal probability integrals for each end of the straight line.

The unsymmetrical curve in Fig. 5(B) is the sum of unsymmetrical dead zones with gains like those indicated in equation 17.

$$\frac{\phi_{io}}{\phi_{ii}}=\frac{m_1}{2}(1-\Phi_0)+\frac{m_2-m_1}{2}(1-\Phi_1)+ \frac{m_3-m_2}{2}(1-\Phi_2)+\ldots+ \frac{m_a}{2}(1-\Phi_0)+\frac{m_b-m_a}{2}(1-\Phi_a)+ \frac{m_c-m_b}{2}(1-\Phi_b)+\ldots \qquad (21)$$

$$=\frac{1}{2}[m_1(\Phi_1-\Phi_0)+m_2(\Phi_2-\Phi_1)+ m_3(\Phi_3-\Phi_2)+\ldots+m_a(\Phi_a-\Phi_0)+ m_b(\Phi_b-\Phi_a)+m_c(\Phi_c-\Phi_b)+\ldots] \qquad (22)$$

Again, the gain is the sum of each slope times the difference of the normal probability integrals for each end of the straight line. It is given a weighting of one half if the slope appears on only one side of the origin, and a weighting of one if it appears symmetrically on both sides. This method permits one to evaluate the gain of a combination of spring-returned backlash and saturation in a control loop, and to specify the range of Gaussian input signals for which the control would be satisfactory.

Output Power of Discontinuous-Slope Nonlinearity

The two kinds of errors due to a clipper are: 1. the desired signal that is not permitted to reach the output; and 2. extra distortion power in the output that should not be there. The magnitudes of these two components will be derived.

To obtain the harmonic power and the power of a nonlinearity or the sum of several nonlinearities, one needs the cross-power between the outputs of two different elements. Consider N_1 and N_2, which are two discontinuous-slope nonlinearities with Gaussian inputs a and b respectively, and with a normalized cross correlation of $\rho_{ab}(\tau)$ between the inputs. The output signals are θ_1 and θ_2 respectively. The CC between the outputs is

$$\psi_{12}=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\theta_1\theta_2\left(\frac{\partial p}{\partial w}\,\frac{\partial p}{\partial x}\right)dw\,dx \qquad (23)$$

where the second probability density is defined by equation 24:

$$\left(\frac{\partial p}{\partial w},\,\frac{\partial p}{\partial x}\right)=\frac{1}{2\pi\sigma_a\sigma_b\sqrt{1-\rho_{ab}^2}}\times \exp\left\{-\frac{1}{2(1-\rho_{ab}^2)}\left[\left(\frac{w}{\sigma_a}\right)^2- 2\rho_{ab}\left(\frac{w}{\sigma_a}\right)\left(\frac{x}{\sigma_b}\right)+\left(\frac{x}{\sigma_b}\right)^2\right]\right\} \qquad (24)$$

Let

$$z=x/\sigma_b \qquad (25)$$

and

$$u=\frac{\left(\dfrac{w}{\sigma_a}-\rho_{ab}^2\right)}{\sqrt{2}\sqrt{1-\rho_{ab}^2}} \qquad (26)$$

Substituting this change of variable into equation 24, and then substituting the density into equation 23, causes the CC to become

$$\psi_{12}=\frac{\sigma_a\sigma_b}{\sqrt{2}\pi}\int_{-\infty}^{+\infty}\frac{\theta_2}{\sigma_b}\,\epsilon^{-\frac{z^2}{2}}\times dz\int_{-\infty}^{+\infty}\frac{\theta_1}{\sigma_a}\,\epsilon^{-u^2}\,du \qquad (27)$$

If N_1 is a symmetrical clipper, θ_1, is w between $-u'$ of $(-c_1-\rho z/\sqrt{2})/\sqrt{1-\rho^2}$, and $+u''$ of $(+c_1-\rho z/\sqrt{2})/\sqrt{1-\rho^2}$, where c_1 is $k_1/\sigma a\sqrt{2}$. The term θ_1 is $-k_1$ below the lower limit, and $+k_1$ above the upper limit.

$$\int_{-\infty}^{+\infty}\frac{\theta_1}{\sigma_a}\,\epsilon^{-u^2}\,du=\int_{-u'}^{+u''}(\sqrt{2}\sqrt{1-\rho^2}\,u+ \rho z)\epsilon^{-u^2}\,du+\int_{-\infty}^{u'}\frac{-k_1}{\sigma_a}\epsilon^{-u^2}\,du+ \int_{+u''}^{+\infty}\frac{k_1}{\sigma_a}\,\epsilon^{-u^2}\,du \qquad (28)$$

$$=\frac{+\sqrt{1-\rho^2}}{\sqrt{2}}(\epsilon^{-(-u')^2}-\epsilon^{-(u'')^2})+ \frac{\rho z\sqrt{\pi}}{2}[\Phi(u'')-\Phi(-u')]+ c_1\sqrt{\frac{\pi}{2}}[2-\Phi(u'')+\Phi(-u')] \qquad (29)$$

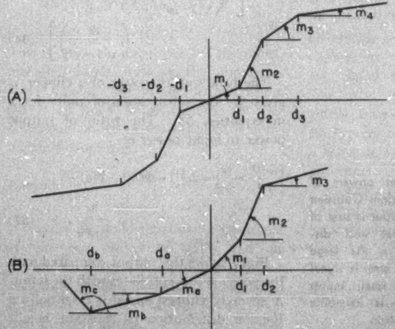


Fig. 5. General nonlinearity represented by straight-line segments. A—Symmetrical about origin. B—Unsymmetrical. Slopes m can be negative, and nonlinearity can have several maxima and minima

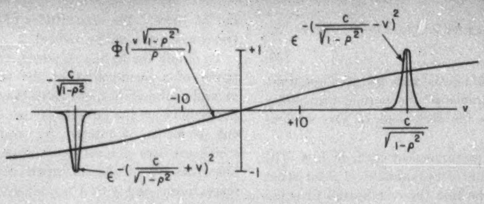


Fig. 6. Components whose product must be integrated to obtain output autocorrelation of discontinuous-slope nonlinearity

The result of the first integration in the foregoing may be substituted into equation 27. Each term must be integrated by parts, and there will be one integral left over. For the particular case when the clipping levels are identical, and a and b are the same, this will yield the output autocorrelation. Divided by the input power, it is

$$\frac{\psi_{oo}(\tau)}{\sigma_i^2}=2c^2[1-\Phi(c)]+\left[\Phi\left(\frac{c(1+\rho)}{\sqrt{1-\rho^2}}\right)+\Phi\left(\frac{c(1-\rho)}{\sqrt{1-\rho^2}}\right)\right]\times\left[+\left(\frac{\rho}{2}+c^2\right)\Phi(c)-\frac{3\rho c}{2\sqrt{\pi}}\epsilon^{-c^2}\right]+\left[\Phi\left(\frac{c(1-\rho)}{\sqrt{1-\rho^2}}\right)-\Phi\left(\frac{c(1+\rho)}{\sqrt{1-\rho^2}}\right)\right]\times\left[\frac{c(1-\rho^2)}{\sqrt{\pi}}\epsilon^{-c^2}\right]-\frac{\rho c^2}{\sqrt{2}\sqrt{1-\rho^2}}\times\int_{-\infty}^{k/\sigma}(\epsilon^{-(u'')^2}-\epsilon^{-(u')^2})\Phi\left(\frac{z}{\sqrt{2}}\right)dz-\frac{\rho c^2}{\sqrt{2\pi}\sqrt{1-\rho^2}}\int_{-k/\sigma}^{+\infty}(\epsilon^{-(u'')^2}-\epsilon^{(-u')^2})\Phi\left(\frac{z}{\sqrt{2}}\right)dz+\frac{\rho}{2\sqrt{2\pi}\sqrt{1-\rho^2}}\int_{+k/\sigma}^{k/\sigma}(\epsilon^{-(u'')^2}-\epsilon^{-(u')^2})\Phi\left(\frac{z}{\sqrt{2}}\right)dz \quad (30)$$

The derivation of equation 30 is quite tedious, although straightforward. The integration of the last terms will be shown only for the case when ρ is near unity, i.e., for a calculation of output power.

The variable must be changed to

$$v=\frac{\rho z}{\sqrt{1-\rho^2}\sqrt{2}} \quad (31)$$

$$u'=\frac{1}{\sqrt{1-\rho^2}}+v;\ u''=\frac{1}{\sqrt{1-\rho^2}}-v \quad (32)$$

The last term in the autocorrelation becomes

$$\frac{1}{2\sqrt{\pi}}\int_{-\frac{c}{\sqrt{1-\rho^2}}}^{\frac{c}{\sqrt{1-\rho^2}}}\left[\epsilon^{-\left(\frac{c}{\sqrt{1-\rho^2}}-v\right)^2}-\epsilon^{-\left(\frac{c}{\sqrt{1-\rho^2}}+v\right)^2}\right]\Phi\left(\frac{\sqrt{1-\rho^2}}{\rho}v\right)dv \quad (33)$$

Fig. 6 shows these functions plotted. The integral goes from the center line of one exponential peak to the center line of the other. All of the significant area is concentrated around these lines. The probability integral acts like a constant-weighting function across the peak, and makes both areas positive. Since there is essentially no area contribution in the regions where Φ is changing, Φ may be assumed to be constant at $+\Phi(c/\rho)$ for positive v, and the same negative value for negative v. Removing this constant outside of the integral, changing the sign for the negative integration, and then writing twice the positive integration for the sum of the two equal areas, yields, for the last integral in equation 30

$$-\frac{1}{2}\Phi\left(\frac{c}{\rho}\right)\left[\Phi\left(\frac{c}{\sqrt{1-\rho^2}}-v\right)+\Phi\left(\frac{c}{\sqrt{1-\rho^2}}+v\right)\right]_0^{\frac{c}{\sqrt{1-\rho^2}}}=\Phi\left(\frac{c}{\rho}\right)\left[\Phi\left(\frac{c}{\sqrt{1-\rho^2}}\right)-\frac{1}{2}\Phi\left(\frac{2c}{\sqrt{1-\rho^2}}\right)\right] \quad (34)$$

In a similar manner, the sum of the other two integrals in equation 30 becomes

$$+c^2\Phi\left(\frac{c}{\rho}\right)\left[\Phi\left(\frac{c}{\sqrt{1-\rho^2}}-v\right)+\Phi\left(\frac{c}{\sqrt{1-\rho^2}}+v\right)\right]_{\frac{c}{\sqrt{1-\rho^2}}}^{\frac{\infty}{\sqrt{1-\rho^2}}}=-c^2\Phi\left(\frac{c}{\rho}\right)\Phi\left(\frac{2c}{\sqrt{1-\rho^2}}\right) \quad (35)$$

The normalized output autocorrelation of a symmetrical clipper is therefore

$$\frac{\psi_{oo}(\tau)}{\sigma_i^2}=2c^2[1-\Phi(c)]+\left[\Phi\left(\frac{c(1-\rho)}{\sqrt{1-\rho^2}}\right)+\Phi\left(\frac{c(1+\rho)}{\sqrt{1-\rho^2}}\right)\right]\times\left[\left(\frac{\rho}{2}+c^2\right)\Phi(c)-\frac{3\rho c}{2\sqrt{\pi}}\epsilon^{-c^2}\right]+\left[\Phi\left(\frac{c(1-\rho)}{\sqrt{1-\rho^2}}\right)-\Phi\left(\frac{c(1+\rho)}{\sqrt{1-\rho^2}}\right)\right]\times\left[\frac{c(1-\rho^2)}{\sqrt{\pi}}\epsilon^{-c^2}\right]+\Phi\left(\frac{c}{\rho}\right)\times\left[\Phi\left(\frac{c}{\sqrt{1-\rho^2}}\right)-\left(\frac{1}{2}+c^2\right)\times\left(\frac{2c}{\sqrt{1-\rho^2}}\right)\right] \quad (36)$$

The total output power of a clipper is given by the substitution of unity for ρ in equation 36. The ratio of output power to input power is

$$\frac{\psi_{oo}(0)}{\sigma_i^2}=\frac{\sigma_o^2}{\sigma_i^2}=2c^2[1-\Phi(c)]+\Phi(c)-\frac{3c}{2\sqrt{\pi}}\epsilon^{-c^2} \quad (37)$$

Fig. 7 shows this output compared with the power which correlates with the input. A severely clipped signal is essentially a Poisson distribution of constant height

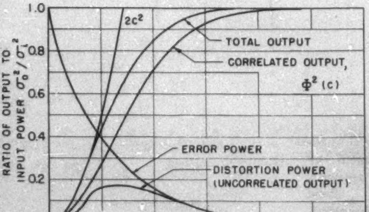


Fig. 7. Output powers of clipper with random Gaussian input. Total output is sum of correlated output and distortion power. At large inputs (small c), error is signal missing. At small inputs (large c), error is harmonic generation

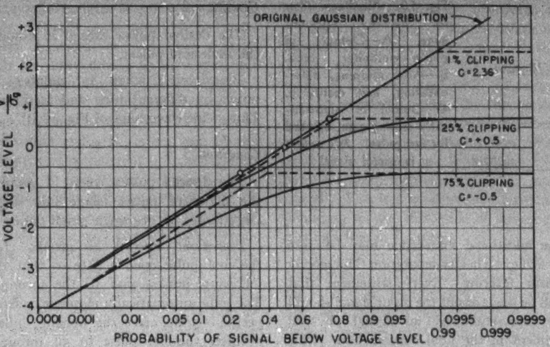


Fig. 8. Measured probability distribution of output of symmetrical clipper with random Gaussian input signal

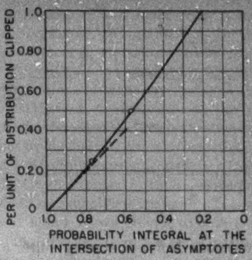


Fig. 9. Determination of spectrum of Gaussian signal prior to clipper from probability measurements after clipping. Curve is the termination of clipped distribution

and random crossings. This has a power of k^2 and a power gain of $2c^2$. The first term in equation 37 represents this severely clipped component.

A good approximation to the total output power is a Φ^2 curve with the abscissa shifted to the left.

$$\sigma_o^2=\psi_{oo}(0)\cong\sigma_i^2\Phi^2\left(c\sqrt{\frac{\pi}{2}}\right) \tag{38}$$

If clipping occurs on only one side, the output power is

$$\sigma_i^2\left[\frac{1}{2}+\frac{1}{2}\Phi^2\left(c\sqrt{\frac{\pi}{2}}\right)\right] \tag{39}$$

The distortion power, or uncorrelated output power, is the total output power in equation 39 for symmetrical clipping minus the correlated power, $\sigma^2\Phi^2(c)$, from equation 14.

$$\sigma_h^2=\psi_{hh}(0)=\sigma_i^2\left[(2c^2+\Phi(c))(1-\Phi(c))-\frac{3c}{2\sqrt{\pi}}\epsilon^{-c^2}\right]\cong\sigma_i^2\left[\Phi^2\left(\sqrt{\frac{\pi}{2}}c\right)-\Phi^2(c)\right] \tag{40}$$

This is, at most, about 18% of the input, but the power lost by the clipper is much greater. From the shapes of the curves, a good approximation for the distortion power in the output for symmetrical clipping is

$$\sigma_h^2\cong\sigma^2\frac{3c}{4\sqrt{\pi}}\epsilon^{-c^2} \tag{41}$$

Nearly all real components in actual systems have saturation levels, and one wishes to design for minimum cost or minimum weight, consistent with small distortion. There is no significant reduction in error power for increases in the clipping level above c of 2. This should be used to select components for high-accuracy proportional systems. If weight or cost increments are quite important, c should be small. A rule of thumb for typical electromechanical designs is a c of $\sqrt{2}$, or a clipping level of twice the rms input.

Non-Gaussian Input to Discontinuous-Slope Nonlinearity

A non-Gaussian input should be represented as being tailored by the same kind of unit as that which it supplies. For clippers or dead zone, one should approximate the input by a Gaussian signal passing through a combination of clippers and dead zone. Then, combine the signal operators with the actual equipment to obtain the over-all operator.

To measure the presence or absence of clipping, the actual probability distribution should be measured. This is done by constructing a clipper which passes all signals above a certain level, and a cumulative timer which runs whenever the signal exceeds the clipping level. The ratio of the timer reading to the actual time is the probability integral. Each point on the integral curve requires a separate run, or an additional clipper and timer.

Plot the measured data on probability integral paper as in Fig. 8. If the curve is a straight line, the distribution is Gaussian. If it is asymptotic to a voltage level, as in Fig. 8, the signal has been clipped on one side at that level. The clipped distribution lies to the right of the original distribution in Fig. 8 at all points. The horizontal value at any point is equal to the value on the original distribution, times the reciprocal of the per-unit clipping level. The slope at low probabilities matches the original distribution. The curved distribution may be moved back to the left until it becomes a terminated portion of the straight-line original distribution. The clipping level is the observed maximum level on the probability plot.

The per-unit clipping may be read directly in Fig. 9, after the asymptote is drawn, in Fig. 8, and the probability at the asymptote intersection read. One should locate the termination point of the original distribution at the clipping level, and a probability of one minus the per-unit clipping. The original distribution line is parallel to the sloping portion of the measured distribution. The original Gaussian rms value σ_y is available from the slope. It is the voltage between probabilities of 0.16 and 0.5. The original and final mean values are also available from the curves. The calculations may be checked, with the use of the previously derived equation 39.

Symmetrical clipping on both sides can be handled in exactly the same manner. The basic clipped curves are derived from the original straight line by applying the reciprocal of the clipping on each side as a correction factor to the deviation of the probability from 0.5 on that side.

This graphical method can also be used for the detection of tailoring by a dead zone, in which case a straight-line segment is removed from the middle of the Guassian distribution. The result of fitting the remaining segments together is that the ends of the curve bend to a greater slope, in the opposite direction from the clipped distribution.

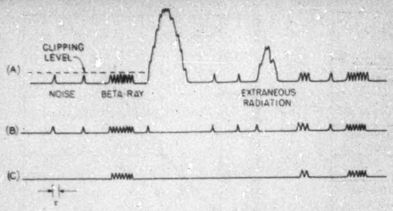


Fig. 10. Nonlinear filter for beta-ray thickness gage. A—Output of photo-multiplier tube. B—After clipping, differentiating, and rectifying. C—Output of automatically closing gate, opened by first pulse. All output is useful signal

Nonlinear Noise-Reducing Filter

One is able to use the criterion of minimum error power to design the best linear filter to separate information from noise, or to design the best linear predictor to compensate for an unalterable linear member. By changing the criterion to minimum error time, and the unalterable element to a nonlinear amplifier, one can obtain a nonlinear race-brake type of design.

In the previous sections, changes in a spectrum due to a nonlinearity have been obtained analytically, but whether these changes are improvements or not has not been considered.

Linear filters are used to separate information from noise when the only thing that is known is the power spectrum. If the amplitude distributions of the information and noise are different, or if the wave forms are different, or if anything else is known besides the power spectrum, the additional knowledge may be used to achieve a better separation than merely linear filtering. Every statement that can be made about the input signal should be represented by an actual physical element in the final construction. This will be illustrated by the following two examples.

A detector for very weak beta rays in the presence of extraneous strong beta radiation could use a scintillation crystal counter and a very long Lucite pipe to carry the light flashes to a photomultiplier tube. Thermal noise on the photocathode would produce individual small pulses. The undesired high-energy radiation in the scintillator would produce high-amplitude pulses. For each low-energy beta ray, light reflections in the long Lucite pipe would produce a train of small pulses, as shown in Fig. 10. Extraneous, undesired radiation and noise pulses can produce both large bursts of great energy and small individual pulses. A circuit can be built to discard most of the unwanted pulses, greatly improving the signal-to-noise ratio.

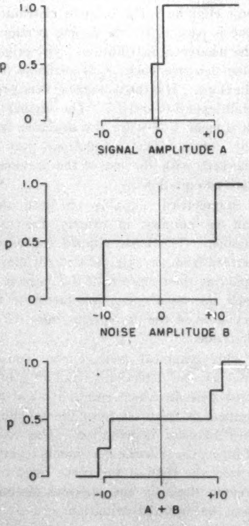


Fig. 11 (left). Amplitude probability integral distribution of signal and noise

Derivation of the Circuit

Condition 1

The number of desired signals is small; the number of undesired signals is large.

Procedure: Clip photomultiplier output at peak of desired signals. This removes most of the undesired energy.

Condition 2

The desired pulses exceed in number the undesired pulses.

Procedure: Change to number of pulses by differentiating and rectifying. This yields waveshape of Fig. 10(B).

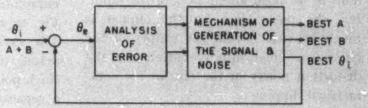
Condition 3

The desired pulses come in groups; the undesired pulses merge into single large pulses, or are single small pulses.

Procedure: Record only groups. Let the first pulse open a gate so that the following pulses pass to the recorder. The gate automatically closes in a short period of time, after the opening pulse, or after the last pulse to go through, so that no individual pulses reach the recorder.

The recorder receives signal shown in Fig. 10(C). This is a remarkable signal-to-noise improvement achieved entirely with nonlinear filtering. The power spectra of the desired and undesired signals were probably very similar. Only the amplitude probability distributions were different.

Another type of nonlinear filter is illustrated by Figs. 11 and 12. The input is a mixture of the two severely clipped Gaussian components in Fig. 11. The mechanism of the origin of the signal was a large amplitude Gaussian distribution which had been severely clipped at plus and minus one volt, so that it had 50% probability of each of these two values, and a negligible probability of any other value. The noise was also a Gaussian distribution which had been clipped at plus and minus 10 volts. The power spectra after clipping were identical, so that no separation of the signal from the noise was possible by linear means. The noise however, was 40 decibels greater than the signal. The signal can be completely separated from the noise by nonlinear means.



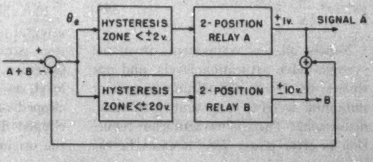


Fig. 12 (right). Nonlinear filter to separate signal and noise in Fig. 11. Filter makes a perfect separation of two components with identical power spectra

A nonlinear filter is similar to a linear filter in that it has a model of the mechanism of generation of the two signals, with the sum fed back to the input and forced to equal the input with a high-gain error amplifier. In addition, there is a block to analyze the error. The model mechanism in Fig. 12 is the sum of the outputs of two different polarized 2-position relays. The *A* relay delivers ±1 volt, and the *B* relay delivers ±10 volts.

If the error exceeds 2 volts, the *B* relay should change, and therefore its hysteresis band is made greater than ±2 volts. If the error is exactly 2 volts, the *A* relay should change and therefore its hysteresis band is less than ±2 volts. The *B* relay will first reproduce the noise input; then the *A* relay will reproduce the signal input. This nonlinear filter will always deliver a correct separation of the two input signals.

It is beyond the scope of this paper to consider nonlinear filters in general, but the method in Fig. 12 can be extended to separating signals with continuous smooth probability distributions. If the two signals each have finite values of probability density at the same amplitudes, then usually they cannot be separated completely without error.

Summary

A nondynamic nonlinearity delivers a power, a part of which is correlated with the input signal, and part of which is distortion, or harmonic power. For Gaussian inputs, the ratio of the CC to the input autocorrelation is the transference, or equivalent gain. The transference squared times the input power is the useful output power. The distortion power is not correlated with the input, but is correlated with the distortion power of other nonlinearities, wherever they may be in the system. For example, a clipper and a dead zone in parallel have distortion powers which may cancel. This is a good method for compensating for the distortion of undesired nonlinearities in a system.

The stability of a closed-loop system containing a nonlinearity is a function of the nonlinear transference. This nonlinear transference can be used with L-plane templates to determine the stability as a function of signal level. The power of the error between the output and the input can be used for specifying the minimum acceptable components for a system.

Non-Gaussian inputs should be represented by a Gaussian distribution tailored by the same kinds of nonlinearities as those which the actual system presents. Then the signal and the system components can be combined for analysis purposes.

A signal can be separated from undesirable noise with a linear dynamic filter if the signal and noise have different power spectra as a function of frequency. Signal and noise can be separated by nonlinear filters if they have the same power spectra but different probability distributions. The philosophy of separation is to build a model of the mechanism of generation of the signal from a flat Gaussian source, and another model of the mechanism of generation of the noise from a different flat Gaussian source. The sum of the outputs of these two models is the best statistical approximation to the input, and it is made identically equal to the input by subtracting the sum from the input and using the error to actuate a high-gain error amplifier. (In linear filters, infinite gain amplifiers are used. In nonlinear filters, relays provide infinite gain at the particular signal level which causes the relay to respond.) The output of the infinite gain error amplifier is examined to determine which error components are due to the signal and which to the noise. These components are separated with a nonlinear function, or with the relay hysteresis, and are used to generate the best guesses for the two flat Gaussian sources which supply the models.

References

1. STATISTICAL SPECTRAL OUTPUT OF POWER LAW NONLINEARITY, Otto J. M. Smith. *AIEE Transactions*, see pp. 535–43 of this issue.

2. *Ibid.*, references 1 through 17.

3. A FIVE-CHANNEL ELECTRONIC ANALOG CORRELATOR, M. J. Levin, A. F. Reintjens. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 8, Jan. 1953, pp. 647–56.

4. THE SPECTRUM OF CLIPPED NOISE, J. H. Van Vleck. *RRL Report no. 51*, Radiation Research Laboratory, New York, N. Y., July 1943.

5. SPECTRAL RESPONSE OF CONTROL SYSTEM CONTAINING ZERO-MEMORY NONLINEARITY TO RANDOM INPUTS (digest), Y. Sawaragi, S. Takahashi. *Mechanical Engineering*, American Society of Mechanical Engineers, New York, N. Y., July 1956, p. 655.

Indialing to PBX Extensions: Application in a Step-by-Step Central Office Area

G. NORMAN SCHLEINKOFER
MEMBER AIEE

UP TO THE PRESENT TIME, the completion of an incoming call to a private branch exchange (PBX) extension telephone has been dependent upon a human attendant. She must determine what extension station is desired and do the proper switching of the call from the incoming trunk to the station. Indialing provides a means whereby the call is dialed directly from the general exchange system to the PBX station without the need for a human attendant, except for transfer and information services.

This paper will discuss the application of indialing in a step-by-step central office area. The headquarters PBX of E. I. du Pont de Nemours and Company (hereinafter referred to as the du Pont PBX) in Wilmington, Del., was chosen for a full-scale field trial. On more than 80% of the incoming calls to this PBX, the extension station was asked for by number. The Olympia central office, which serves this PBX, was readily adaptable for indialing. The customer had requested this type of service in connection with some additions and rearrangements that had become necessary.

Indialing is an extension of the mechanization program of the telephone system. This program began with the introduction of dial equipment, for local use only, early in the century and has progressed to a point where almost all stations are on a dial basis and a large part of these can dial points far beyond the local service area. Dialing of calls within

---

Paper 59-185, recommended by the AIEE Communications Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 10, 1958; made available for printing December 8, 1958.

G. NORMAN SCHLEINKOFER is with The Bell Telephone Company of Pennsylvania, Philadelphia, Pa.

a PBX and, to a limited extent, to the local area outside the PBX was introduced at about the same time.

Without indialing, the attendant requires additional time for each incoming call. This increases the average holding time of every call made to a PBX. The additional time consists of a few seconds for the trunk lamp to be illuminated, several seconds for the attendant to become conscious that a call is waiting and be free to answer it, quite a few more seconds to determine from the caller just what extension is desired, and finally a few seconds to make the connection and start ringing. The total is estimated to average about 30 seconds for the PBX under consideration.

The saving of this number of seconds may appear insignificant when only one call is considered. However, when applied to a large proportion of the calls to PBX's over the system, a considerable saving is made possible in the trunk plant, toll switching equipment, and operating effort of the telephone companies. The customer obtains better incoming service a reduction in the personnel for operating the switchboard, and a smaller monthly charge for the switchboard positions as well as other incidental and intangible advantages. As a by-product of the numbering system that must be used with indialing, it is possible for the customer to adopt better bill checking and accounting procedures.

To provide the advantages of indialing to everyone will require time and money. Time is involved because it is not possible to convert all PBX's for indialing without disrupting the orderly supply of material for growth, mechanization, and replacement of the remainder of the telephone system, because comparatively large amounts of equipment are required for each job. Money must be spent for the establishment of a new numbering scheme for the PBX, as well as for rearrangements of the trunk plant of the entire telephone system, the central office equipment, and the equipment at the PBX. Each of these expenses will be considered separately.

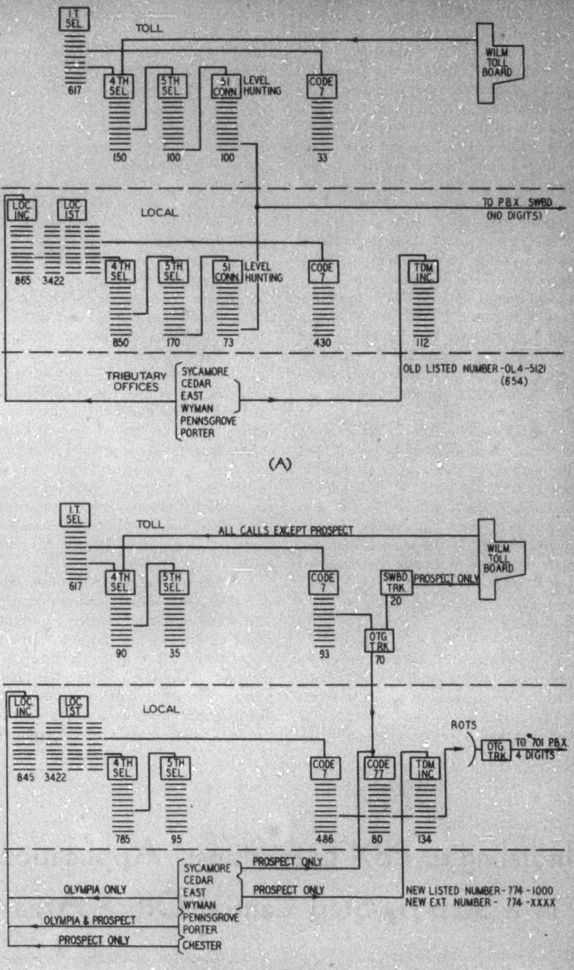

I.T. SEL.
TOLL
617
4TH SEL.
5TH SEL.
51 CONN.
LEVEL HUNTING
CODE 7
WILM. TOLL BOARD
150
100
100
33
TO P.B.X. SWBD. (NO DIGITS)
LOC. INC.
LOC. 1ST
LOCAL
865
3422
4TH SEL.
5TH SEL.
51 CONN.
LEVEL HUNTING
CODE 7
TDM. INC.
850
170
73
430
112
TRIBUTARY OFFICES
SYCAMORE
CEDAR
EAST
WYMAN
PENNSGROVE
PORTER
OLD LISTED NUMBER-OL4-5121 (654)
(A)
I.T. SEL.
TOLL
ALL CALLS EXCEPT PROSPECT
617
4TH SEL.
5TH SEL.
CODE 7
SWBD. TRK.
PROSPECT ONLY
WILM. TOLL BOARD
20
90
35
93
OTG. TRK.
70
LOC. INC.
LOC. 1ST
LOCAL
845
3422
4TH SEL.
5TH SEL.
CODE 7
CODE 77
TDM. INC.
ROTS
OTG. TRK.
TO 701 PBX 4 DIGITS
785
95
486
80
134
SYCAMORE
CEDAR
EAST
WYMAN
PENNSGROVE
PORTER
CHESTER
PROSPECT ONLY
PROSPECT ONLY
OLYMPIA ONLY
OLYMPIA & PROSPECT
PROSPECT ONLY
NEW LISTED NUMBER-774-1000
NEW EXT. NUMBER- 774-XXXX
(B)

Fig. 1. Associated central office equipment, Wilmington, Del.

A—Before du Pont PBX indialing
B—After du Pont PBX indialing

Numbering Considerations

The first step toward indialing is to assign a 7-digit code consisting of two letters and five numerals to each PBX extension station. This 7-digit code must be compatible with the nation-wide numbering plan so that it may be reached from any telephone in the system. In the case of the du Pont PBX, an entire central office code with a possibility of 10,000 telephones was reserved. The code 774 was given the name PRospect 4. To this code was added the 4-digit extension number.

At the present time, the Delaware and eastern Pennsylvania areas are able to dial interchangeably on a 7-digit basis. This limits the assignment of a code to one that has not been pre-empted previously. Since Philadelphia has some panel-type central offices equipped with translator-senders, these limit the codes available to 200. A further limitation in the choice of codes was due to the arrangement of switches in the Olympia central office. Hence only those codes which could be assigned to established selector levels could be used. There was only one code available in the Olympia central office which had not been reserved for other purposes, so that it will not be possible to provide indialing services to other large customers until additional codes have been released from their present reservation.

The number of codes available will increase as the direct distance-dialing program progresses, the translator-senders of the Philadelphia central offices are replaced and it is necessary to install additional equipment in the Olympia central office for other purposes.

After the new code had been selected it was placed in the route information sheets of every central office operator in the entire telephone system, so that it could be reached from any point. New route cards were cut at all points arranged to switch direct distance-dialed calls automatically.

The listed call number to be used required careful consideration and was included in the series assigned to extension stations. If not selected carefully, it will reduce the codes remaining for assignment for other purposes. At the du Pont PBX the number chosen was PR 4-1000. Incoming calls were routed from the first level of the first selectors directly to the switchboard. This left the first level of the local first selectors free to be used for tie trunk, foreign exchange, and other service codes. The remaining levels may all be assigned to stations as required.

Trunking Changes

Since a new central office code was established to provide indialing, it was necessary to furnish this new central office with its complement of incoming trunks somewhat similar to the arrangement for the existing one. A study was made of the incoming traffic and its present routing. This study indicated the need for new groups of trunks from some of the near-by central offices, the augmentation of some of the present trunk groups, and the decrease of other trunk groups.

Since traffic from more distant points would arrive in Wilmington over the same paths as formerly, it was only necessary to establish paths from the toll switchboard and the toll train switches to the new code levels.

Trunking in the Olympia central office was most severely changed, since with indialing all of the traffic for this PBX will now pass through its own central office levels rather than through the Olympia train of switches. As can be seen by comparing Figs. 1(A) and 1(B), some of the equipment formerly used is no longer required and some must be installed just because of the new code.

There are three types of trunks feeding into the new PRospect 4 code:

1. Local trunks from the Olympia, Porter, Penns Grove, and Chester central offices.
2. Trunks from the toll switchboard and distant points, as well as Sycamore.
3. Trunks from points served by the local tandem network (East, Cadar, and Wyman central offices).

These trunks feed into separate groups of switches which accept a varying number of digits. The traffic from the switch groups is then combined and carried to the PBX through rotary out-trunk switches and outgoing repeaters which deliver the last four digits of the extension station number of the incoming selector.

Dial pulse-repeating circuits are used when trunking a call from one step-by-step central office to another and at times within the central office, especially when going from one type of circuit to another. If more than two repeaters are used in tandem, correcting circuits must be used to avoid pulse mutilation. Pulse-correcting circuits are expensive and are hard to maintain and therefore are not used unless another alternative is not available.

Some of the trunk routings were rearranged so that a maximum of only two pulse repeaters were in any one connection. It was also necessary to redesign the originally contemplated transfer circuit so as to eliminate a repeating feature.

Outgoing trunks from the PBX are used in the same manner as formerly. Foreign exchange and tie trunk points are reached by dialing available codes which have been derived from the first and other levels. Until the levels are needed for station purposes, local outgoing trunks are reached by dialing "9" and the attendant, by dialing "0." When these levels are required for station use, the services furnished by them will be transferred to some other available level, perhaps "19" and "10." This arrangement is the same as that usually provided for PBX's without indialing.

Local outgoing and incoming trunks are on a one-way basis, since no economy would result from combining the two groups.

Central Office Rearrangements

A circuit from a dial central office to an ordinary PBX is furnished from the same type of central office equipment that is used for a single telephone. The progress of the call is controlled by dial pulses until it leaves the central office when the signaling is changed to ringing current which actuates the equipment at the telephone or PBX. Because of the design of this equipment, dial pulses may not be passed from the central office to the PBX

However, by transferring some of the functions of the central office to the PBX, circuits can be installed so that dial pulses may be passed over them to operate switches and select a telephone at the PBX. In this way, the call leaves the central office on what otherwise might be considered an intra-office trunk.

Fig. 1(A) shows the arrangement of the central office equipment before indialing. At that time there were 76 PBX trunks from the Olympia step-by-step central office (OL 4-5121) to the 32-position manual switchboard at the PBX. These trunks are reached by means of level hunting connector switches which provided full access to groups of 100 trunks, in contrast to only 10 which may be reached through the usual type of step-by-step switch.

Local and toll calls used separate trains of switches and groups of level hunting connectors. Since the banks were multiplied, each switch has full access to all trunks.

Fig. 1(B) shows the arrangement of the central office equipment after indialing. The numerals under the switch symbols indicate the total number of switched of that type in the Olympia central office. By comparing the numbers as shown in Figs. 1(A) and 1(B), the differences in the equipment requirements may be seen.

The principal changes made in the central office were as follows:

1. An additional group of selector switches (Code "77") was installed.
2. Additional switches were placed in the Code "7" selector group of the toll train.
3. More switches were added to the local incoming selector group.
4. More switches were added to the Code "7" tandem selectors and the tandem incoming selectors.
5. New switchboard trunks from the toll switchboard and additional outgoing pulse repeating trunks from the Code "7" selectors of the toll train to the new Code "77" selectors were installed.
6. New rotary out-trunk switches and dial

pulse repeaters were provided between the central office selectors and the PBX.

7. The level hunting connectors of both the toll and local trains were eliminated.

8. The number of switches in the remaining groups of selectors of the local and toll trains except the local first selectors and incoming toll selectors were reduced.

Since the Olympia central office was equipped with digit-absorbing first selectors, a complete train of selectors to use each digit has not been installed. This type of switch obviated a costly equipment addition when this central office was changed from 5- to 7-digit dialing for local calls. However, the number of codes that can be reached is limited. All the available levels on the present selectors were in use so that a new bank of switches known as Code "77" selectors was installed to permit traffic from certain trunk groups to be routed to the new PR 4 central office code.

There were available in the Olympia central office outgoing repeaters similar to those in service for other purposes which were used to carry the circuits to the PBX. These were required to convert the circuits from 3-wire to 2-wire so as to avoid the use of extra cable pairs and sleeve repeaters.

Rotary out-trunk switches were provided on the trunks leading from the central office to the PBX for the following reasons:

1. To provide trunk efficiency equivalent to that available with level hunting connectors which were formerly in service.

2. It was found to be less costly to provide the rotary out-trunk switches than to install the extra equipment and cable plant for the additional trunks required if rotary out-trunk switches were not used.

3. The calling customer would have a considerably better chance of reaching the PBX without encountering a "false busy."

The local first selector and incoming toll selectors remain the same because substantially the same traffic feeds into them, although leaving via different levels. The differences in the quantities of other switches reflect the changed paths that the traffic must take to reach the new call number.

Most of the central office equipment which was released by the rerouting of the traffic could be readily reused for growth for other customers with the exception of the level hunting connectors.

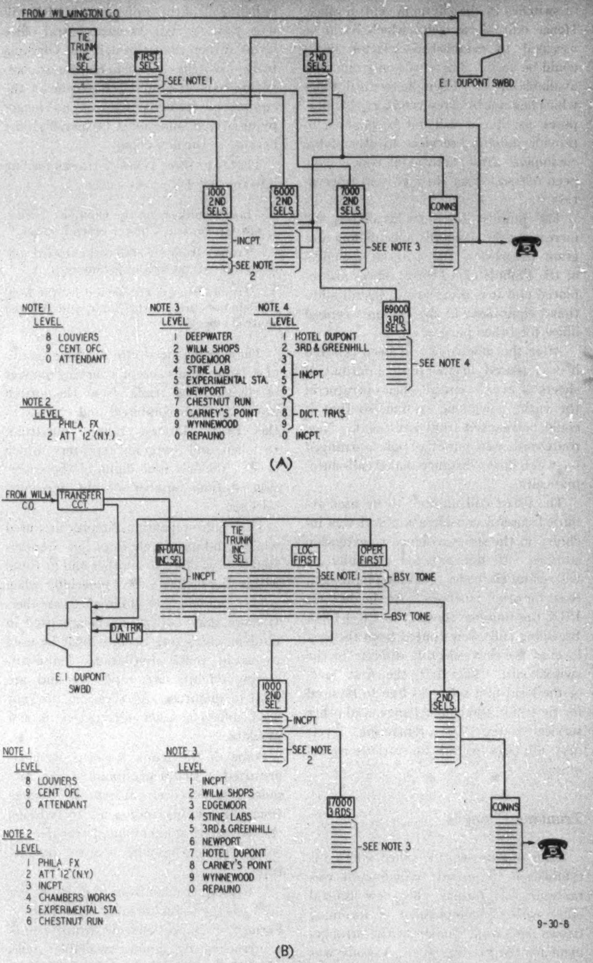

FROM WILMINGTON C.O
TIE TRUNK INC. SEL
FIRST SELS
SEE NOTE 1
2ND SELS
E.I. DUPONT SWBD
1000 2ND SELS
INCPT.
SEE NOTE 2
6000 2ND SELS
7000 2ND SELS
SEE NOTE 3
CONNS
69000 3RD SELS
SEE NOTE 4
NOTE 1
LEVEL
8 LOUVIERS
9 CENT. OFC.
0 ATTENDANT
NOTE 2
LEVEL
1 PHILA. FX
2 ATT '12' (N.Y.)
NOTE 3
LEVEL
1 DEEPWATER
2 WILM. SHOPS
3 EDGEMOOR
4 STINE LAB
5 EXPERIMENTAL STA.
6 NEWPORT
7 CHESTNUT RUN
8 CARNEY'S POINT
9 WYNNEWOOD
0 REPAUNO
NOTE 4
LEVEL
1 HOTEL DUPONT
2 3RD & GREENHILL
3 4 5 6 INCPT.
7 8 9 DICT. TRKS.
0 INCPT.
(A)
FROM WILM. CO.
TRANSFER CCT.
IN-DIAL INC. SEL
TIE TRUNK INC. SEL
LOC. FIRST
OPER FIRST
INCPT.
SEE NOTE 1
BSY. TONE
BSY. TONE
DA TRK UNIT
E.I. DUPONT SWBD
1000 2ND SEL
INCPT.
SEE NOTE 2
2ND SELS
17000 3RDS
SEE NOTE 3
CONNS
NOTE 1
LEVEL
8 LOUVIERS
9 CENT. OFC.
0 ATTENDANT
NOTE 2
LEVEL
1 PHILA. FX
2 ATT '12' (N.Y.)
3 INCPT.
4 CHAMBERS WORKS
5 EXPERIMENTAL STA.
6 CHESTNUT RUN
NOTE 3
LEVEL
1 INCPT.
2 WILM. SHOPS
3 EDGEMOOR
4 STINE LABS
5 3RD & GREENHILL
6 NEWPORT
7 HOTEL DUPONT
8 CARNEY'S POINT
9 WYNNEWOOD
0 REPAUNO
9-30-8
(B)

Fig. 2. E. I. du Pont headquarters PBX

A—Before indialing
B—After indialing

PBX Additions and Changes

Indialing requires that some of the functions of the central office be preformed at the PBX. The two letters and one numeral of the central office code of each extension station are used in the central office for directing the call to the customer's location and only the last four digits are delivered to the first switch in the PBX which directs the call to the extension station. This is compatible with the present method which uses four digits to dial from one station to another within the PBX.

The new equipment required at the PBX consisted of the incoming selectors, the transfer circuits, the trunk circuits for the listed number trunks, an intercepting arrangement, and switchboard jack appearances for these circuits. The traffic from the incoming selectors is routed over common trains of second selectors and connectors with all other types of calls. The switches in these trains were increased to handle the added traffic. This resulted in a considerable saving in equipment over that required if separate trains had been used.

The tariff which applies to PBX equipment in Delaware permits the telephone company to install the number of switches required for a standard grade of service. In most other states the customer pays for the number of switches installed and can choose any grade of service he desires. It is important that all switches associated with indialing be under the control of the telephone company; otherwise, the action of the customer could have a serious effect on telephone service in general.

When indialing was first suggested for the du Pont PBX, it was equipped for 4,900 lines and 32 positions of the latest type of switchboard, having a 7-panel station multiple. However, the use of four of these positions had been suspended. Because of indialing, six additional positions were disconnected. Although the use of a total of ten positions was discontinued, only six were removed, leaving the other four for expected growth in other traffic.

The reduction in the number of positions due to indialing is difficult to evaluate. Many of the positions are used exclusively for tie trunk and foreign exchange traffic, and the number used for answering incoming calls varies with the need. All positions are not staffed continuously but are programed to follow the load. In this way, a particular position may be used for different purposes at different times. Although it is expected that 90% of the incoming calls will be in-dialed, this does not mean that the incoming load will be reduced by 90%. The remaining 10% of the calls are expected to be those that require the most effort on the part of the attendant. In addition, the attendant must still perform any transfer and information services required.

Because more than average space was occupied by the trunk multiple, additional station jacks would have been beyond the reasonable reach of the attendant. It was necessary, therefore, to resort to some other method of completing calls from the switchboard to the telephone at the extension station. Attendant dialing of all these calls was the alternative used and required the replacement of all the station multiple with switching equipment.

At the same time that indialing was installed, 300 additional station lines were added for growth, bringing the total to 5,200. This required using the seventh level for station assignment. The seventh level was formerly used for tie trunk codes. Other codes derived from the first level of the first selectors were used for this purpose, involving a rearrangement of the trunking.

Although 76 incoming trunks had been in service before indialing, the load on them had grown to a point that more than standard delays were encountered during the busy hour. To bring the delay factor back to standard and to care for the additional business generated by the 300 new stations required that 97 trunks be placed in service. Each of these trunks has a jack appearance on the switchboard to permit the transfer of calls.

A separate complement of jacks was provided in the switchboard to answer listed number calls (PR 4-1000). However, since these calls were routed through the selector banks it was necessary to provide only enough jacks to carry the traffic rather than install one for each trunk.

To be able to charge properly for the indialed calls and not charge for those that reached an unused terminal, a recorded message type of intercepting service was installed. It was not feasible to have the attendant give this information, since a call could be dialed to a known unused terminal and, when the attendant answered, a message that should be paid for could be given to her without being charged to the calling telephone. Calls from outside the PBX, as well as those dialed within it reach the same unused terminals. This made it necessary to develop wording which would give understandable information to both types of calls in one message.

Fig. 2(A) shows the switching equipment at the PBX before indialing and Fig. 2(B) shows it afterward.

Attendant Dialing

Attendant dialing at this PBX was a a necessity because of space limitations in the face of the switchboard. However, the introduction of attendant dialing should be considered at the same time as indialing. It is not economical to provide attendant dialing prior to indialing because large amount of switchboard equipment needed to handle the calls that will be indialed subsequently. If attendant dialing is provided after indialing, some false work work may be done in installing additional multiple, which may have to be removed later. For large PBX's with indialing the attendant dialing equipment is considerably less expensive than multiple, especially when common trains of switches are used.

In choosing between 2-, 3-, or 4-digit attendant dialing, four digits were adopted. This resulted in a smaller number of items of equipment and a better arrangement of jacks. Two- or 3 digit dialing would require a varying number of jacks in a large number of groups, which could not be satisfactorily fitted into the 20-jacks-per-strip arrangement of the switchboard.

The removal of the station multiple made available much needed space into which the trunk multiple was expanded. This expandion was due not only to the jacks needed for attendant dialing trunks that replaced the station multiple but also for increased requirements in the number of central office trunks, tie trunks, attendant trunks, and other services that have jack appearances.

Transfer Circuit

In order to give full PBX service required at the headquarters of the du Pont Company, it was necessary to provide for the transfer of calls from one extension station to another. Since this service is not used in central offices, nor in PBX's up to the present time, the local customer services engineers were obliged to develop a suitable transfer circuit.

This circuit must perform the following functions:

1. Pass the incoming call to the selectors with a minimum of interference with the dial pulses.
2. Notify the central office equipment when the called station answers, and also when it disconnects.
3. Recognize the difference between flashing for a transfer and a permanent disconnection.
4. Cause a lamp to light on the switchboard when a transfer request is received.
5. Prevent any transfer request from reaching the preceding equipment so as to avoid a premature disconnection.
6. Provide a 3-way connection between the calling and called parties and the PBX attendant.
7. Release the local switching equipment between the transfer circuit and the called telephone after the transfer information has been given to the PBX attendant and the original called party has hung up.
8. Establish a new talking path between the incoming circuit and the switchboard.
9. Allow any number of subsequent transers under the control of the switchboard cords.
10. Hold the trunk busy until the calling party disconnects, but permit the release of the local switches if the called party disconnects first.

11. Restore to normal when the distant called party disconnects.

12. Deny access to the trunk by the attendant before the first transfer request is made to maintain secrecy and to avoid interference with the initial progress of the call.

All of these functions were combined in a relay unit which was manufactured by the Western Electric Company and installed in each indialing trunk circuit.

The transfer circuit was arranged so that it passed dial pulses directly into the indialing selectors. When the called party answered, a repeating coil was inserted in the circuit to separate the central office and PBX circuits. This separation permitted a delay in passing supervision to the calling central office. This delay was necessary to prevent a transfer flash from appearing as a disconnect signal.

Conclusions

The indialing equipment was officially placed in service at the opening of business on July 7, 1958. There was no "cutover" as such, because practically all the circuits had been cut through for almost a week previously and had been used for test calls and by others who had the necessary information.

The existing trunks from the Olympia call number were not immediately disconnected but were left in service to avoid an excessive intercepting load at the central office and to reduce the necessity of having the calling customer place two calls if he either was not informed of the change in number or neglected to use it. Any caller who used the old number was informed of the new one by the PBX attendant before completing the call so that he would know the proper dialing procedure for subsequent calls.

Almost immediately, indialed calls amounted to over 80% of the total and soon built up to about 90%.

Indialing, together with attendant dialing and the transfer feature, operated to the entire satisfaction of this customer. However, this is not to be taken as an indication that indialing would be satisfactory for every PBX. Different problems of transmission, supervision, numbering, dialing, load characteristics, and trunking must be solved in each case.

Basic Concepts of Multidimensional Space Filters

GABRIEL KRON
FELLOW AIEE

Synopsis: The conventional electric network, consisting of linear or one-dimensional *RLC* (resistance-inductance-capacitance) elements, may be generalized by adding to it planar-, cubic-, etc., up to k-dimensional *RLC* elements. The resultant geometrical configuration of 0-, 1-, 2-. . . to k-dimensional spatial networks interconnected into one structure is called, in combinatorial topology, a "polyhedron." Its electric-network realization may be called a "wave model." The electromagnetic structure formed by interconnecting two or more wave models into one network, will be called a "multidimensional space filter." The latter is expected to be employed eventually for the theoretical and perhaps practical solution of such physical problems for which the conventional linear (one-dimensional) electric filter is used today, on a much more elaborate multidimensional scale, of course. A single polyhedron is expected to be used in the numerical solution of simpler multidimensional problems of the calculus of finite differences, having nonuniform space intervals.

This paper illustrates the establishment of the basic tensors of a space filter by means of a tetrahedron. In particular, it is shown that if the space filter is torn apart structurally into a set of isolated polyhedra, and if each polyhedron is torn apart dimensionally into a set of isolated q-networks (each with q-dimensional elements), the theories and equations of a linear 1-network are applicable to each q-network separately. It is, however, necessary that the equivalent linear network should be considered as an "orthogonal" network having nonsingular (square) connection tensors C and A. Each q-network also has a complete set of junction-pair and mesh equations properly coupled (the complete set of equations is needed when the isolated q-networks are interconnected). The incidence matrices arising after interconnection are also given. The steps for propagating an electromagnetic wave through an isolated q-network are shown and the wave vectors are considered to define geometrically a qth order "divided difference." Simplified algebraic diagrams are also developed. During the solution process these diagrams are to be used in place of q-dimensional graphs.

The process of interconnecting the component q networks into a wave model and several of the latter into a space filter, utilizes the entire apparatus of the method of tearing.[1] A series connection is shown in the paper.

Two Storehouses of Information

In trying to analyze and solve extremely complicated physical, economic, or other systems with a very large number of variables, the author is gradually evolving a new philosophy of attack. Throughout the analysis and solution of the system he utilizes simultaneously two storehouses of information: 1. the mathematical equations of the system, and 2. a topological model of the system, or the system itself.

The model (or the system) is assumed to consist of two component parts: 1. an underlying "dead" topological structure, which represents the independent variables and structural parameters in the form of multiply connected spaces, and 2. a superimposed "live" algebraic structure, which represents the dependent variables and nonstructural parameters in the form of electromagnetic and other physical vectors propagating across the various spaces and between the various polyhedra.

Both geometry and physics associated with the model are continuously put to use to direct the mathematical unfolding of the solution process. At every intermediate step there exists a 3-pronged alteration upon, and a three fold exchange of information between the mathematics, geometry, and physics of the system. The triangular manipulation and flow of information continues in unison until the final solution of the problem has been reached.

THE ORTHOGONAL POINT OF VIEW

To apply linear networks for the piece-by-piece analysis and solution of large-scale nonelectrical problems, a comprehensive orthogonal point of view was developed.[1,2] The unified theory represents a combination of the mesh and junction-pair points of view. The over-

---

Paper 59-228, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted October 27, 1958; made available for printing December 3, 1958.

GABRIEL KRON is with the General Electric Company, Schenectady, N. Y.

all view brought to light the simultaneous existence of four types of currents and four types of voltages in the same linear network.

Contrary to recent trends, the author has consistently refused to drag in the concepts of junction points (0-space) and planes (2-space) as constituent or auxiliary parts of a linear network. By stopping at the upper and lower boundaries (at the meshes and junction pairs), and not extending beyond them, a self-consistent topological structure and algebraic diagram could be constructed for a linear network. As a result, all linear concepts remain valid, first of all for a single higher-dimensional network. Moreover, the existence of self-consistent building blocks makes it possible to interconnect a whole sequence of multidimensional networks into one logical structure, a polyhedron (or wave model), with the same facility, as several subdivisions of a linear network are interconnected into one structure. The interconnection of wave models into a space filter follows the steps for constructing supernetworks out of already solved networks.

ANTICIPATED FIELDS OF APPLICATION

Generally speaking, the use of a polyhedron appears to the author as a novel method for solving a class of problems in which the large number of independent variables and parameters are nonlinear. (The dependent variables are still linear as long as the volume elements of the polyhedron are Euclidean spaces. Of course, the volume elements may be assumed to be non-Euclidean.) The most obvious and simplest application of a polyhedron, or rather, a wave model, is to generalize the various formulas of the calculus of finite differences for nonuniform intervals assumed in a k-dimensional Euclidean space. This paper restricts itself to generalizing the concept of divided differences of various orders. They are expected to be used for interpolation, curve fitting, smoothing, numerical differentiation, and integration, and for other problems with which the calculus of finite differences deals.

Partial differential equations, integral equations, and integropartial differential equations require too many net points in more than two or three dimensions, if the net points are regularly spaced. However, a comparatively few points spaced at random can adequately represent very large dimensional spaces. Thus it is hoped that the space-filter approach will eventually open up new avenues for solving systematically advanced types of physical problems described by simultaneous equations with a very large number of independent variables and parameters. The space filter is expected to consider during the solution process not only the unknown function, but also all its higher derivatives near the given points, all simultaneously, as the several polyhedra will be properly staggered within the confines of the enveloping k space.

It is anticipated that advanced problems in numerical analysis (Fourier transform, power density spectra), in operations research (quadratic and higher programing), in economics, etc., will also offer opportunities for the use of space filters. The study of diffraction and scattering of electromagnetic waves in asymmetrical crystal and molecular structures would only be a first step into an ever-widening field of physical application for multidimensional space-filter concepts.

THE PROCESS OF TEARING

In solving problems of linear networks[1] the networks were usually torn into subdivisions only when the number of cuts was far less than the number of original variables. In that case, the intersection matrix z_4 to be inverted had far fewer rows and columns than the inverted matrix Z of the original network. The polyhedron offers an example where the intersection network, constructed by tearing the system apart dimensionwise, has practically as many rows and columns as the original network. However, the factorized form of the inverse is now triple-diagonal, while the true inverse of the original network is a full matrix.

Moreover, the hidden variables liberated at the cuts represent now the main informations sought, such as the regression coefficients (in curve fitting), or the displacements of atoms (in diffraction studies). In linear networks the hidden variables are needed for auxiliary use only and are not of primary interest.

The analysis and solution of wave models and space filters in a piecewise manner appears to be unavoidable from either an organizational or a computational point of view. Since the process of tearing and its accompanying topological concepts are unknown to both engineers and physicists, this fact alone may explain the reason for the complete absence of polyhedra and space filters in the solution of physical or numerical problems. The occasional use of polyhedra in the design of experiments, or in molecular packing, etc., has no relation to the present study since the polyhedra are not energized. This paper endows the polyhedron with a superimposed electromagnetic wave; that is, here a "dead" polyhedron is turned into a "live" one.

SIX BASIC STEPS IN A "PRIMITIVE" POLYHEDRON

No matter for what analytical problem the space filter is employed, the first step in its application is to tear the space filter apart structurewise into isolated polyhedra, then tear each polyhedron apart dimensionwise into isolated q-networks and perform as many operations on the isolated q-networks as possible. This paper deals with six of the basic operations that are to be performed on each q-network (the "primitive" polyhedron) in any problem. All these manipulations are repetition, or straightforward generalization, of those performed on a torn but conventional linear network. The six basic steps are:

1. A sequence of 0-, 1-, 2-, to k-dimensional networks is constructed between n vertices defined by the independent variables.

2. Each q-network is torn apart further into its constituent volume elements, the primitive q-network.

3. A diagonal $Z_{\alpha\beta}{}^{(q)}$ or $Y_{(q)}{}^{\alpha\beta}$ matrix, the "metric tensor," is set up for each q-network, representing the square of the q-dimensional volume of the component elements. (A full matrix may be set up if so desired, containing also all the direction cosines.)

4. Each q-network is interconnected separately by establishing a nonsingular square "connection tensor" $C_{\alpha'}{}^{\alpha(q)}$. Its inverse $A^q = C_t^{q-1}$ is also calculated.

5. Each nonsingular connection matrix C and its inverse A is partitioned into two rectangular matrices, the "boundary operators" C_m, C_j, and A^m, A^j. (Out of the latter a set of rectangular "incidence matrices," M, and reciprocal incidence matrices M^{-1} may be set up. These rectangular matrices show the relations between two neighboring q-networks, and serve as a check on the correctness of the boundary operators.)

6. By assuming the dependent variables $y_\alpha{}^{(p)}$ (one or p sets) as bound magnetic poles $\dot{p}_{m0}{}^{(b)} = \dot{B}_0$ (alternating at say 60 cycles) placed on the vertices of the 0-network, an electromagnetic wave is allowed to spread across the open-circuited space filter. The transversal waves with $\dot{e}_\alpha$, $\dot{b}_m$, h^α, and $\dot{d}_\alpha$ vectors, as well as the longitudinal waves with E_m, $\dot{B}_\alpha$, H^α, and $\dot{D}_\alpha$ (or $\dot{p}_m{}^{(f)}$, $\dot{p}_m{}^{(b)}$, $\dot{\rho}_{(f)}{}^{e}$, $\dot{\rho}_{(b)}{}^{e}$) vectors, appearing on each q-network, represent generalized divided differences of qth order, or rather, their covariant and contravariant forms. No back-reaction exists during this open-circuited propagation.

As the manipulation of each q-network involves further succession of pointwise or functional tearings, the practical solution of any space-filter problem involves thus a hierarchy of "tearing within tearing." The author is confident that the basic theory of such multiple tearings

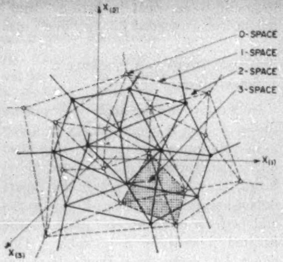


Fig. 1. Three-dimensional space filter formed by two polyhedra

has been satisfactorily covered in his publications of diakoptics.[1,2]

The connection matrix C^q of each q-network consists of +1, −1, and 0 elements. Their inverse calculation is rather simple and can be performed by longhand. Also all impedance matrices $Z_{\alpha\beta}{}^{(q)}$ and $Y_{(q)}{}^{\alpha\beta}$ belong to the isolated networks only and are diagonal. That is, the establishment of the sequence of divided differences will involve only the products of matrices with vectors (for one dependent variable).

Expressed in another manner, in simple problems the polyhedron is left open-circuited. The wave propagation is directed by the engineer and not by the structure itself. In more advanced studies the entire gamut of conventional wave propagations will appear in generalized forms. The networks will be connected into a polyhedron.

Constructing the Sequence of Networks

TEARING INTO ISOLATED NETWORKS

In order to establish the basic equations of state of a system of k multidimensional networks, the same steps will be followed as are used by the author in the analysis of other complex systems:

1. Tear the complex into isolated q-networks. Each network includes its own upper and lower boundaries, but not those of its neighbors.
2. Set up the equations of state of each q-network separately, as if the other q-networks were nonexistent.
3. Interconnect the q-networks and their equations of state with the aid of a connection tensor C representing their topological interrelations.

Fortunately each two neighboring q-networks have the same number of upper or lower boundaries. They are also similarly situated in space. Hence their interconnection into one system will be simple and painless. The connection matrix C consists of two equal unit matrices once the boundary axes have been matched.

0-DIMENSIONAL (POINT) NETWORK

To each value of the dependent variable y_α (with n components) belongs one value of each of the k independent variables $x_{(k)}{}^\alpha$. These k values of $x_{(k)}{}^\alpha$ are assumed as one point in a k-dimensional Euclidean space. Altogether there are n such points or vertices. The free index α is assumed to vary from one to n.

These n vertices define the 0-dimensional network. Its lower boundaries are the subnetworks (one, in the present case). Its upper boundaries are formed by the junction pairs (their number is $n-1$).

ONE-DIMENSIONAL (LINEAR) NETWORK

The next step is to erect, between the n points, a linear structure that covers the k-dimensional space with a network of one-dimensional lines or reference axes. Since these axes cannot be the conventional orthogonal straight lines, the aim is to approximate the latter as closely as possible (Fig. 1).

If all distances between sets of two points have been calculated, then: 1. the points inside the k-space are connected with about $2k$ points that are nearest; 2. the points at corners of the k-space are connected to only about k neighboring points; and 3. the points along the edges have $k+1$ neighbors, and so on.

The square of the length of each branch will be called "dielectric reluctance" $\epsilon_{\alpha\beta}{}^1$ (or impedance $Z_{\alpha\beta}{}^1$). The linear network is a dielectric network, at least in the interpolation problem.

q-DIMENSIONAL NETWORKS

The next step is to connect neighboring branches with planes (Fig. 1). Each plane will lie between three (or more) vertices. The aim is again to approximate a set of orthogonal planes connecting neighboring lines. The square of a triangular area will be called "magnetic reluctance" $\mu_{\alpha\beta}{}^2$ (or impedance $Z_{\alpha\beta}{}^2$). The planar network is a magnetic network.

Next, the neighboring triangles are connected with tetrahedrons, defined by four (or more) points, to form a 3-dimensional (cubic) reference network (Fig. 1). Then a set of five points defines a 4-dimensional network and in general a set of $q+1$ vertices defines a q-dimensional network.

The square of a q-dimensional volume element defines an impedance $Z^{(q)}$ (i.e. $\epsilon_{\alpha\beta}{}^{(q)}$ or $\mu_{\alpha\beta}{}^{(q)}$). For each primitive network the impedances are arranged along the diagonal of a matrix $Z_{\alpha\beta}{}^{(q)}$. Its inverse is $Y_{(q)}{}^{\alpha\beta}$ (i.e., $\epsilon_{(q)}{}^{\alpha\beta}$ or $\mu_{(q)}{}^{\alpha\beta}$). In general, the even-dimensional networks are magnetic, the odd-dimensional networks are dielectric.

q-DIMENSIONAL JUNCTION PAIRS AND MESHES

Each q-network will be organized in the same manner as a linear network has

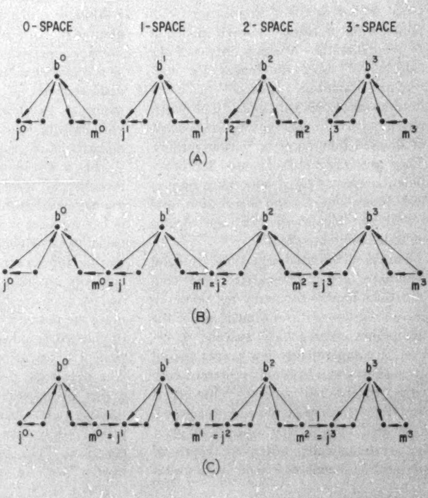


Fig. 2. Sequence of q spaces and their boundary spaces

A—Before interconnection
B—After interconnection
C—Separating joint boundary spaces by unit arrows

been organized. The steps will follow those developed in references 1 and 2.

With a linear 1-network having $b^{(1)}$ branches, two types of boundary concepts are associated. One of the concepts implies elements of a 0-network, the other concept suggests those of a 2-network. These two types of concepts (so-called "cycles") are

1. $j^{(1)}$ junction pairs. Each branch (or several branches) is bounded by two vertices (elements of a 0-network). The vertices do not belong to the 1-network.

2. $m^{(1)}$ meshes. A mesh forms the boundary of a plane (element of a 2-network). The planes do not belong to the 1-network.

The relation between the independent number of cycles (Betti numbers) and the number of branches is

$$b^{(1)}=j^{(1)}+m^{(1)} \tag{1}$$

Boundaries and Coboundaries

With a q network having $b^{(q)}$ component volume elements (branches), two types of boundary concepts (cycles) are associated, one implying $q-1$, the other $q+1$-dimensional volume elements. These two types of cycles are

1. $j^{(q)}$ junction pairs or boundaries. Each q volume of the q network is bounded by a $q-1$-dimensional surface. The latter is not part of the q-network.

2. $m^{(q)}$ meshes or coboundaries. A mesh (several q-surfaces) forms the boundary of a $(q+1)$ volume. The latter is again not part of the q-network.

The relation between the independent number of cycles (Betti numbers) and q-volumes is

$$b^{(q)}=j^{(q)}+m^{(q)} \tag{2A}$$

Interconnection of q-Networks

Each q-network (and its two types of boundaries) is established independently of the others. When the boundaries are interconnected into a sequence of networks, the upper boundary (m) of a $q-1$ space unites with the lower boundary (j) of a q-space to form one boundary. That is, the number of neighboring boundary and coboundary elements is the same

$$m^{(q-1)}=j^{(q)} \tag{2B}$$

Thus one portion of the connection matrix consists of two equal unit matrices with $j^{(q)}$ diagonal unities

$$A_q^{q-1}=C_q^{q-1}=\begin{array}{c|c} & c_q^{q-1} \\ m^{q-1} & 1 \\ j^q & 1 \end{array} \tag{2C}$$

Each q-network may be connected to its upper or lower neighbor either in shunt, or in series, or in some other manner. (Similarly the several polyhedra may be connected to each other in a large variety of ways and may also be inductively coupled to each other.) The manner of interconnection depends on the problem at hand and the nature of known and unknown quantities. The entire apparatus of diakoptics[1] with its mesh or junction intersection network concepts is needed to interconnect the isolated q-networks into a polyhedron and to interconnect the isolated polyhedra into various types of space filters.

Orthogonal Networks

Orthogonal Linear Networks

In order to understand the necessity for the construction of square, or rather, nonsingular, connection matrices C^q, it is necessary to sidestep and review the self-consistent orthogonal point of view of a linear network. The theory is also self-sufficient, as it must and does explain all characteristics of a linear network without dragging in extraneous features, such as junction points or planes.

Conventional theories analyze a linear electric circuit (after Maxwell) in two different ways:

1. As a mesh network, containing $m^{(1)}$ meshes only and described by a set of $m^{(1)}$ equations:

$$e_m=z_{mm}i^m \tag{3A}$$

2. As a junction network, containing $j^{(1)}$ junction pairs only, and described by a set of $j^{(1)}$ equations:

$$I^j=Y^{jj}E_j \tag{3B}$$

(Maxwell used a single datum point and not arbitrary junction pairs.)

In order to use electric-circuit theories for the analysis and solution of nonelectrical problems (or for more advanced electrical problems) the author has developed, in references 1 and 2, a more general orthogonal theory that includes the mesh and junction-pair theories as special cases. A linear network is assumed to possess simultaneously both meshes and junction pairs, and thus it is possible to write, for any linear network, as many equations as there are branches, $b^{(1)}$.

$$\begin{array}{l|l} e_m=z_{mm}i^m+z_{mj}I^j & I^j=Y^{jj}E_j+Y_{jm}e_m \\ E_j=z_{jm}i^m+z_{jj}I^j & i^m=Y^{mj}E_j+Y^{mm}e_m \end{array} \tag{3C}$$

That is, each network possesses simultaneously two types of currents and voltages:

1. Solenoidal (mesh) vectors i^m, e_m (denoted by lower-case letters).

2. Lamellar (jp) vectors I^m, E_m (denoted by capital letters).

More general quantities and equations have also been introduced[1,2] but they are not needed for the immediate problem.

Orthogonal q-Networks

The same orthogonal point of view will now be applied to each multidimensional q-network separately, with some modification. The main change is that the networks will conduct not electric currents, but electromagnetic waves (and charges). In particular: 1. all odd-dimensional spaces (1, 3, 5. . .) conduct two types of dielectric quantities: (a). solenoidal $\dot{d}^\alpha$, e_α, and (b). lamellar $\dot{D}^\alpha$, E^α; and 2. all even-dimensional spaces (0, 2, 4. . .) conduct two types of magnetic quantities: (a). solenoidal h^α, b_α, and (b). lamellar H^α, $\dot{B}_\alpha$.

Accordingly the impedance matrices will be denoted by $\epsilon_{\alpha\beta}$, $\epsilon^{\alpha\beta}$ and $\mu_{\alpha\beta}$, $\mu^{\alpha\beta}$ respectively, but often the $Z_{\alpha\beta}$ and $Y^{\alpha\beta}$ notation will also be used.

Note: The author suggests that in routine calculations it is sufficient to use only linear network terminology and notation e, I, z, and Y for the entire system. The number of dimensions may be denoted by an index.

The physically correct notation and terminology must be retained, however, to present the new concepts in their proper setting.

Topological Representation of Charges

To the author at least, it is a very remarkable fact that all lamellar (capital letter) quantities have to be identified with electric charges ρ^e and magnetic poles ρ_m, in order to satisfy Maxwell's equations. In particular, let: 1. $\rho_{(b)}$ represent "bound" charges, and 2. $\rho_{(f)}$ represent "free" charges.

In that case the following identification can be made:

$$\begin{array}{l|l} \dot{D}^\alpha\equiv\dot{\rho}_{(b)}{}^e & H^\alpha\equiv\dot{\rho}_{(f)}{}^e \\ E_\alpha\equiv\dot{\rho}_m{}^{(f)} & \dot{B}_\alpha\equiv\dot{\rho}_m{}^{(b)} \end{array} \tag{4A}$$

In considering the incidence matrices M as differential operators, this identifi-

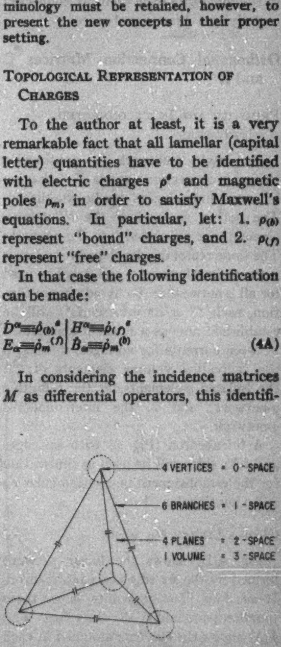


Fig. 3. Tetrahedron

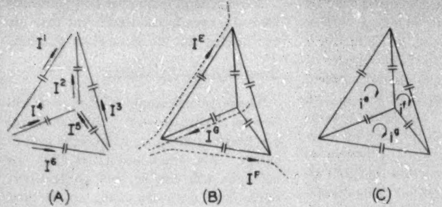


Fig. 4. Linear networks

A—Primitive 1-network
B—Junction pairs
C—Meshes

cation satisfies Maxwell's equation in all dimensions.

$$\text{div}(\dot{D}^{\alpha}+\dot{d}^{\alpha})=H^{\alpha}=\dot{\rho}_{(f)}^{e}$$
$$\text{div}(\dot{B}^{\alpha}+\dot{b}^{\alpha})=E_{\alpha}=\dot{\rho}_{m}^{(f)} \qquad (4B)$$

This identification (in multiply-connected spaces) of lamellar flux lines with point charges is in line with the modern trend to explain the physical existence of charges and particles by assuming the underlying 4-dimensional physical space as multiply connected.[3] In that case the charges and particles are no longer assumed to be singularities in a simply connected space, but are flux lines trapped in the topology of the multiply connected underlying 4-dimensional space.

Orthogonal Connection Matrices C and A

Primitive and Interconnected q-Networks

For each q-network a nonsingular (square) connection matrix $C_{\alpha}{}^{\alpha}$, having $b^{(q)}$ rows and columns, can be established. The matrix shows how the $b^{(q)}$ volume elements are connected into a q-network. The same reasoning developed for a linear network in references 1 and 3 will be used for all q-networks. To avoid circumlocution, each C^q(or its inverse $A_t{}^q$) will be established now as a relation maintained between currents (or voltages) flowing in the individual volume elements of: 1. the isolated q-elements, or the primitive q-network, and 2. the interconnected q-network.

A tetrahedron (Fig. 3) with six edges and four sides will be used as illustration for the establishment of a nonsingular C^1 and a nonsingular C^2.

C^1 of a Linear Network

Fig. 4(B) shows the linear network formed by the six edges of a tetrahedron. It has four junction points, hence three junction pairs and 6−3=3 meshes. Fig. 4(A) shows the currents assumed in each isolated branch. Equating the currents flowing in each branch before and after interconnection, the following relation exists:

$$I^1=C^1(I^{1'}+i^{1'}) \qquad (5A)$$

$$\begin{array}{c} \\ 1\\2\\3\\4\\5\\6 \end{array}\begin{array}{c} \\ \begin{bmatrix} I^1\\I^2\\I^3\\I^4\\I^5\\I^6 \end{bmatrix} \end{array} \begin{array}{c} \\ = \end{array} \begin{array}{c} \begin{array}{ccc|ccc} I^E & I^F & I^G & i^e & i^f & i^g \end{array} \\ \left[\begin{array}{ccc|ccc} 1 & & & 1 & & \\ & & & -1 & 1 & \\ & & & & -1 & \\ & & -1 & -1 & & 1 \\ & & & & 1 & -1 \\ & 1 & & & & -1 \end{array}\right] \end{array}$$

$$\begin{array}{c} [I^{1'} \quad i^{1'}] \\ =[C_j{}^1 \quad C_m{}^1] \end{array} \qquad (5B)$$

A vertical partition splits the square C^1 into two rectangular matrices. The mesh method of Maxwell utilizes $C_m{}^1$ only.

The inverse of C^{-1} is

$$A_{1t}=\begin{array}{c|cccccc|} & 1 & 2 & 3 & 4 & 5 & 6 \\ E & 1 & 1 & 1 & & & \\ F & & & -1 & & -1 & 1 \\ G & & 1 & & -1 & -1 & \\ \hline e & & -1 & -1 & & & \\ f & & & -1 & & & \\ g & & & -1 & & -1 & \end{array}=\begin{bmatrix} A_{1t}{}^j \\ A_{1t}{}^m \end{bmatrix} \qquad (5C)$$

A horizontal line partitions the square into two rectangular matrices. Only $A_{1t}{}^j$ is utilized by the junction method of Maxwell.

C^2 of a Planar Network

Fig. 5(A) shows the four planes of the tetrahedron. Fig. 5(B) shows the three independent meshes that bound three of the planes, and Fig. 5(C) shows the single closed surface formed by four planes. Thus 4=3+1. The closed surface is represented by one current I^s only.

Equating the currents flowing in each plane before and after interconnection, the following nonsingular relation exists:

$$I^2=C^2(i^{2'}+I^{2'}) \qquad (6A)$$

$$\begin{bmatrix} I^a\\I^b\\I^c\\I^d \end{bmatrix} = \begin{array}{c} \begin{array}{ccc|c} i^e & i^f & i^g & I^s \end{array} \\ \left[\begin{array}{ccc|c} 1 & & & 1 \\ & 1 & & 1 \\ & & 1 & 1 \\ & & & -1 \end{array}\right] \end{array} = \begin{array}{c} [i^{2'} \quad I^{2'}] \\ {[C_j{}^2 \quad C_m{}^2]} \end{array} \qquad (6B)$$

Now the previous meshes play the role of junction pairs (boundaries) and the closed surface plays the role of meshes (coboundaries).

The inverse of C^2 is

$$A_{2t}=\begin{array}{c|cccc|} & a & b & c & d \\ e & 1 & & & 1 \\ f & & 1 & & 1 \\ g & & & 1 & 1 \\ \hline s & & & & -1 \end{array}=\begin{bmatrix} A_{2t}{}^j \\ A_{2t}{}^m \end{bmatrix} \qquad (6C)$$

Incidence Matrices M and M^{-1}

Algebraic Diagram of a Linear Network

It has been shown in reference 1 that, in a linear network, the two square connection matrices C and A, as well as their four partitioned matrices (boundary operators) can be arranged into an algebraic diagram (vector diagram) shown in Fig. 6(A). Each arrow represents a matrix and the sum of two arrows represents the product of two matrices. (In taking sums, the directions go against the arrows by convention.) A set of unit matrices are also shown by arrows.

As the algebraic diagrams will get rather involved later on, it will be convenient to show only the four rectangular matrices C_m, C_j, and A^m, A^j [Fig. 6(B)], supplemented by two unit matrices. It will be understood from now on that the four rectangular matrices (boundary operators) are derived from the square connection matrix C and its inverse A by partitioning the latter.

The complete algebraic diagram of a linear network includes the partial and total impedance and admittance matrices Z and Y, as well as the partial and total connection matrices C and A. A rather complete diagram is shown in Fig. 7(A). The diagram shows some (but not all) of the vectors and matrices that appear in equation 3(C). A still more simplified form, to be used in most studies, is shown

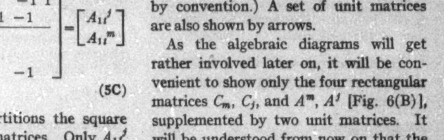


Fig. 5. Planar network

A—Primitive 2-network
B—Junction pairs, meshes
C—Meshes, closed surfaces

. 78
1
59

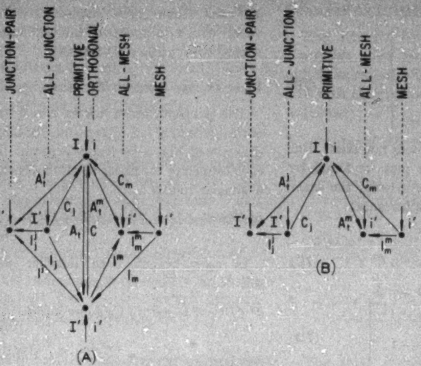

JUNCTION-PAIR
ALL-JUNCTION
PRIMITIVE
ORTHOGONAL
ALL-MESH
MESH
(A)
(B)

Fig. 6 (above). Current spaces of a linear network

A—Complete
B—Simplified form

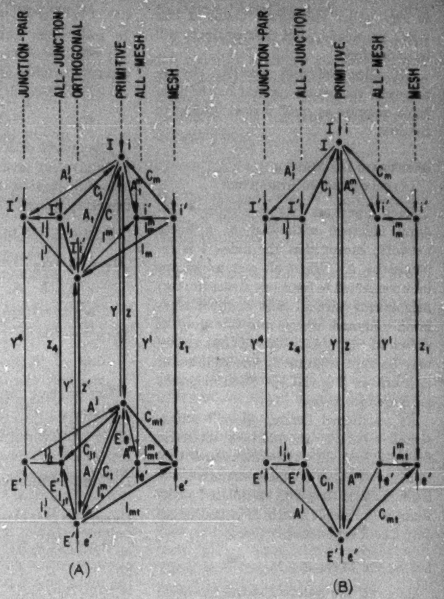

JUNCTION-PAIR
ALL-JUNCTION
ORTHOGONAL
PRIMITIVE
ALL-MESH
MESH
(A)
(B)

Fig. 7 (right). Algebraic diagram of linear network

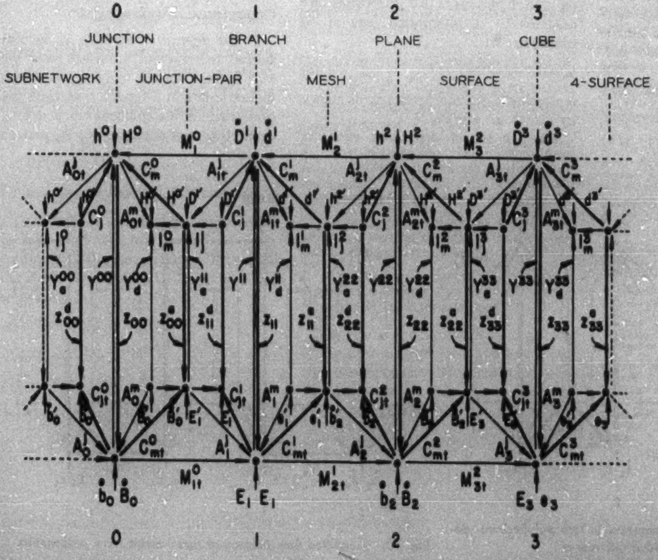

0
1
2
3
JUNCTION
BRANCH
PLANE
CUBE
SUBNETWORK
JUNCTION-PAIR
MESH
SURFACE
4-SURFACE
0
1
2
3

Fig. 8. Simplified algebraic diagram of polyhedron

in Fig. 7(B). The square C and A matrices are missing, but their existence will be understood.

All solenoidal and lamellar electromagnetic vectors also appear on the algebraic diagram as small arrows leading to the dots.

ALGEBRAIC DIAGRAM OF INTERCONNECTED STRUCTURE

Each q-network has an algebraic diagram identical with that of a linear network, except that the index 1 is replaced by q. When all q-networks are interconnected to form one structure, two neighboring vertical lines on their algebraic diagrams merge into one line, as shown in Fig. 8(A). The vertical arrows may be kept separate by the addition of unit arrows [Fig. 2(C)]. Both diagrams are simplified forms.

As mentioned before, all odd-dimensional networks are built of dielectric material and all even-dimensional networks are built of magnetic materials. It is emphasized that with an isolated q-network, no incidence matrix M is associated, only a connection matrix C or A.

INCIDENCE MATRICES M AND M^{-1}

Between two q-networks the incidence matrices M appear as horizontal lines. They are rectangular matrices, not square, and contain only $+1$, -1, and 0 elements. They show how the volume elements of two networks with different dimensions are interconnected. The incidence matrices are found as products of two boundary operators (partial connection matrices).

From Fig. 7(B) the two middle slant vectors give

$$M_2{}^1 = C_m{}^1 A_{2t}{}^J \quad | \quad M_{q+1}{}^q = C_m{}^q (A_{q+1}{}^J)_t \tag{7A}$$

$$M_2{}^1 = \begin{array}{c|cccc} & a & b & c & d \\ \hline 1 & 1 & & & 1 \\ 2 & -1 & 1 & & \\ 3 & & -1 & & -1 \\ 4 & -1 & & 1 & \\ 5 & & 1 & -1 & \\ 6 & & & -1 & -1 \end{array} \tag{7B}$$

The inverse of $M_2{}^1$ does not exist. But an equivalent inverse may be found by following (against the arrows) the inner two slant vectors. For instance:

$$(M_2{}^1)^{-1} = M_1{}^2 = C_j{}^2 A_{1t}{}^m$$

$$(M_{q+1}{}^q)^{-1} = M_q{}^{q+1} = C_j{}^{q+1} A_{qt}{}^m \tag{7C}$$

$$M_1{}^2 = \begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline a & & -1 & -1 & & & \\ b & & & -1 & & & \\ c & & & -1 & & -1 & \\ d & & & & & & \end{array} \tag{7D}$$

The product $M_2{}^1 M_1{}^2$ is not unity and the notation M^{-1} is probably not appropriate (only convenient).

Divided Differences

DEPENDENT VARIABLES

The stage is now set to introduce the electromagnetic waves. That is, hitherto the underlying topological structure was established and the next step is to superimpose an algebraic structure over it.

The set of dependent variables y_α (with n components) will be assumed as a set of bound magnetic poles $\dot{\rho}_m{}^{(b)o} = \dot{B}_\alpha{}^b$ placed on the 0-vertices. All quantities are assumed to be alternating (say at 60 cycles) so that

$$y_\alpha = \dot{B}_\alpha{}^0 = d\rho_m{}^{(b)}/d^\prime t \tag{8}$$

represents a generated voltage, as shown by a short arrow in Fig. 9.

When p set of dependent variables $y_\alpha{}^{(p)}$ exist, then $\dot{B}_\alpha{}^{0(p)}$ represents not a vector with n components, but p different vectors (that is, an $n \times p$ matrix).

FIRST-ORDER DIVIDED DIFFERENCES

The propagation of waves across the entire structure is shown on the flow diagram of Fig. 9 by heavy arrows. On each branch of the linear network appears a lamellar voltage:

$$\dot{\rho}_{m1}{}^{(f)} = E_1 = A_1{}^J C_{mt}{}^0 \dot{B}_0 = M_1{}^0 \dot{B}_0 = \operatorname{div} \dot{B}_0 \tag{9A}$$

In each branch both lamellar and solenoidal dielectric fluxes flow

$$\dot{d}^1 + \dot{D}^1 = \epsilon^{11} E_1 = \epsilon^{11} A_1{}^J C_m{}^0 \dot{B}_0 = \epsilon^{11} M_1{}^0 \dot{B}_0 \tag{9B}$$

representing the first-order divided differences (or rather, their contravariant form).

The total stored dielectric energy in the entire linear network is equal to the sum-of-squares of all first-order divided differences

$$\eta_1 = \Sigma E_1 \epsilon^{11} E_1 = \Sigma(\dot{d}^1 + \dot{D}^1)\epsilon_{11}(\dot{D}^1 + \dot{d}^1)$$
$$= \Sigma(\dot{d}^1 + \dot{D}^1) E_1 = \left(\frac{y_n - y_{n-1}}{x_n - x_{n-1}}\right)^2 \tag{9C}$$

(Strictly speaking, each divided difference is represented not by a flux, but by the square root of a stored energy. The conventional divided differences are not components of a tensor. More will be said about this point in a later paper.)

GEOMETRICAL INTERPRETATION

If the dependent variable y_α is considered to represent a surface over the $x_{(k)}{}^\alpha$ space, then the line integral of each dielectric flux in a branch represents the projection of the gradient upon a line, connecting two points of the y_α surface

Vol.
Pt. 1
1959

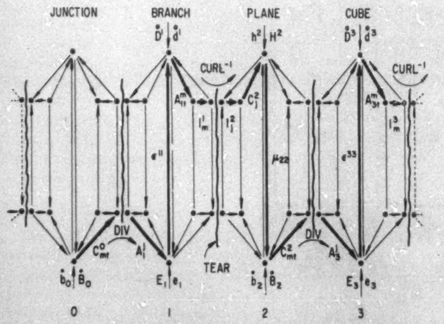


Fig. 9. Flow diagram of wave propagation in torn polyhedron; determination of divided differences

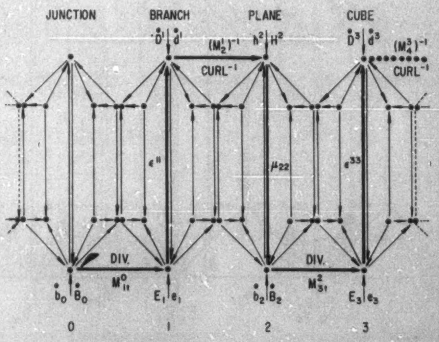


Fig. 10. Simplified flow diagram of open-circuit wave propagation

(not the slope of the surface itself). The higher order divided differences relate, in a complicated manner, the higher order divergences and curls ("exterior" derivatives[4,5]) with the planes and cubes of the polyhedron. For example:

$$\dot{b}_{q+1}+\dot{B}_{q+1}=\text{curl }(e_q+E_q) \tag{10}$$

The divided differences of the present paper are generalizations of those used in the calculus of finite differences, and should not be confused with derivatives.

SECOND-ORDER DIVIDED DIFFERENCES

On each plane (or on some of the planes only) of the planar network appears, according to the flow diagram

$$\begin{aligned} h^2&=C_{ji}^2A_1^J(\dot{d}^1+\dot{D}^1)\\ &=(M_2^1)^{-1}(\dot{d}^1+\dot{D}^1) \qquad \text{(11A)}\\ &=\text{curl}^{-1}(\dot{d}^1+\dot{D}^1)\end{aligned}$$

It should be noted that the path across the boundaries of the two q-networks is directed now against the incidence matrix M_2^1, requiring its inverse, which is, however, nonexistent (since M_2^1 is rectangular). This is the reason for the need of orthogonal network concepts: to open up otherwise nonexistent pathways of propagation.

Since the planes are assumed to be constructed of magnetic materials, the magnetic flux lines (lamellar and solenoidal) appearing across them are given by

$$\begin{aligned}\dot{b}_2+\dot{B}_2=\mu_{22}h^2&=\mu_{22}C_{ji}^2A_1^J(\dot{d}^1+\dot{D}^1)\\ &=\mu_{22}(M_2^1)^{-1}(\dot{d}^1+\dot{D}^1) \qquad \text{(11B)}\end{aligned}$$

representing the second-order divided differences (or rather their covariant form). Again the total stored magnetic energy represents their sum-of-squares:

$$\begin{aligned}\eta_2&=\Sigma h^2\mu_{22}h^2=\Sigma(\dot{b}_2+\dot{B}_2)\mu^{22}(\dot{B}_2+\dot{b}_2)\\ &=\Sigma h^2(\dot{B}_2+\dot{b}_2) \qquad \text{(11C)}\end{aligned}$$

HIGHER ORDER DIVIDED DIFFERENCES

One cycle of propagation has thus far been accomplished and the second cycle continues in the same manner with only the dimensions changed. In general, the steps are

$$\begin{aligned} &E_q=(M_q^{q-1})_t(\dot{B}_{q-1}+\dot{b}_{q-1})\\ &\dot{d}^q+\dot{D}^q=\epsilon^{qq}E_q=Y^{qq}E_q \qquad \text{(12)}\\ &h^{q+1}=(M_{q+1}^q)^{-1}(\dot{d}^q+\dot{D}^q)\\ &\dot{b}_{q+1}+\dot{B}_{q+1}=\mu_{q+1,q+1}h^{q+1}=Z_{q+1,q+1}h^{q+1}\end{aligned}$$

SUMMARY OF DIVIDED DIFFERENCES

The tensorial divided differences of successive order are calculated with the aid of the rectangular boundary operators and with the previous divided differences as follows:

1. $\Delta y/\Delta x=\dot{d}^1+\dot{D}^1=\epsilon^{11}A_1\dot{D}^JC_m^0\dot{B}_0$ $=\epsilon^{11}M_1^0\dot{B}_0$
2. $\Delta^2y/\Delta x^2=\dot{b}_2+\dot{B}_2$ $=\mu_{22}C_{ji}^2A_1^J(\dot{d}^1+{}^1\dot{D})$ $=\mu_{22}(M_2^1)^{-1}(\dot{d}^1+\dot{D}^1)$
3. $\Delta^3y/\Delta x^3=\dot{d}^3+\dot{D}^3$ $=\epsilon^{33}A_3^JC_m^2(\dot{b}_2+\dot{B}_2)$ $=\epsilon_{33}M_3^2(\dot{b}_2+\dot{B}_2)$ (13)
4. $\Delta^4y/\Delta x^4=\dot{b}_4+\dot{B}_4$ $=\mu_{44}C_{ji}^4A_3^J(\dot{d}^3+\dot{D}^3)$ $=\mu_{44}(M_4^3)^{-1}(\dot{d}^3+\dot{D}^3)$

The total stored energy in each q-network is equal to the sum-of-squares of the qth order divided differences.

With one set of dependent variables y_α, each divided difference is a vector. Their calculation involves products of a matrix with a vector. With several sets of $y_\alpha^{(p)}$ each divided difference is an $n\times p$ matrix (p different vectors).

All covariant ϵ_{pp} or μ_{qq} may be denoted as Z_{pp} or Z_{qq} and all contravariant ϵ^{pp} or μ^{qq} may be denoted as Y^{pp} or Y^{qq}. The incidence matrices M_p^q and their inverse $(M_p^q)^{-1}$ represent either divergence or curl operators. A simplified flow diagram for the calculation of the divided differences in terms of M's and Z's is shown in Fig. 10. It is emphasized that different problems require slightly different definitions for the divided differences.

CYCLES OF CREATION AND ANNIHILATION

The following interesting facts should be noted concerning the nature of propagation of waves.

1. In passing to the next higher dimension in a horizontal direction the transverse (solenoidal) waves or the longitudinal (lamellar) waves alternately disappear.
2. In propagating across the material network in a vertical direction a missing portion of the wave reappears.
3. After each two dimensions the cycle of annihilation and creation of an entire electromagnetic wave repeats itself.

The reason for the cyclic zig zag shape of the flow diagram in Fig. 9 is the fact that two successive differentiations cancel a wave entirely. Familiar examples of such cancellations are

$$\text{div curl}\Phi=0 \mid \text{curl grad}\Phi=0 \tag{14A}$$

A generalization of the divergence and curl concepts in simply connected spaces is the concept of "exterior" differentiation arising in multiply connected spaces[4,5] used here. Two successive exterior differentiations give

$$dd\Phi=0 \mid M_q^{q-1}M_{q+1}^q=0 \tag{14B}$$

(In the present example the differential operator d is represented by a rectangular incidence matrix M.)

THE METRICS OF DE RHAM AND HODGE

With the artifice of introducing material constants $\epsilon_{\alpha\beta}$ and $\mu_{\alpha\beta}$ as metric tensors and zigzagging between covariant and contravariant spaces, the entire sequence of networks has been populated with transverse and longitudinal electromagnetic waves.

Unfortunately the geometrical or physical analogs of the concepts of impedance tensor $Z_{\alpha\beta}$ and admittance tensor $Y^{\alpha\beta}$ do not appear in the works of De Rham[4] or Hodge[5] nor in Misner and Wheeler[3] in spite of the introduction of metrics into their spaces (generalized polyhedron). These writers use the metric tensor for introducing isomorphism between the primary and dual wave quantities that arise in relativity physics and not for introducing ϵ and μ into the underlying space. Of course, these writers do not introduce any cyclicly repetitive wave propagation either. The latter is an artifice needed for a nonuniform calculus of finite differences and is not likely to arise in nature or in geometry.

References

1. DIAKOPTICS—THE PIECEWISE SOLUTION OF LARGE-SCALE SYSTEMS, G. Kron. *The Electrical Journal*, London, England, June 7, 1957, issue and following.

2. TENSOR ANALYSIS OF NETWORKS (book), G. Kron. John Wiley & Sons, Inc., New York, N. Y., 1939.

3. CLASSICAL PHYSICS AS GEOMETRY, C. W. Misner, J. A. Wheeler. *Annales de Physique*, Paris, France, vol. 2, no. 6, Dec. 1957, pp. 525–603.

4. VARIETES DIFFERENTIABLES (book), G. de Rham. Hermann, Paris, France, 1955.

5. THE THEORY AND APPLICATION OF HARMONIC INTEGRALS (book), W. V. D. Hodge. Cambridge University Press, Cambridge, England, 1952.

6. A GENERALIZATION OF THE CALCULUS OF FINITE DIFFERENCES TO NONUNIFORMLY SPACED VARIABLES, Gabriel Kron. *AIEE Transactions*, pt. I (*Communication and Electronics*) vol. 77, Sept. 1958, pp. 539–44.

Inductance of A-C Magnets from Simple Models

J. F. H. DOUGLAS
FELLOW AIEE

R. J. VOITH
NONMEMBER AIEE

THE INDUCTANCE of an a-c magnet is a function of its coil as well as its magnetic structure. Hence, if the total equivalent permeance of the magnetic field is known, the inductance can be obtained from the equation

$$L = N^2 \mathcal{P}_{eq}(10^{-8}) \qquad (1)$$

where

L = inductance, henrys
N = series turns in the coil
$\mathcal{P}_{eq}$ = equivalent permeance, lines/ampere

Because computing the permeance of even very simple fields is at best a laborious process, the permeance is generally most easily obtained from experimental tests on models. This paper describes a method in which the field of an a-c magnet is divided into a particular set of hypothetical components which can be simulated electrically with simple models. The permeance of each component field is obtained from a model which suits the component field, and the several permeances so obtained are superimposed to yield the permeance of the original field.

Since there are, in general, fields which satisfy Laplace's equations and those which do not, two kinds of models are used. A component field which has a potential function at all points may be simulated by an electroconducting analog model, while a component field which has no potential function must be simulated with a model having a coil and iron parts, or else found by a close mathematical approximation. Both kinds of models are used in this work. However, in stating that the several component fields may be found from models, the mathematical solution of a particular component field is not excluded as a less desirable method. Occasionally a particular component field can be easily computed with sufficient accuracy and one such example is given in Appendix I.

In order to furnish a check on the accuracy of the method and the applicability of the models described in this paper, the specifications of a prototype magnet were used in constructing the models. Tests were then conducted on the prototype magnet which furnished the basis of comparison.

Component Fields of an A-C Magnet

According to the principle of superposition, a magnetic field may be separated into several hypothetical components for purposes of analysis if no saturation is present. It was assumed that the magnet selected for this work operated at flux densities low enough to satisfy this requirement. Fig. 1 shows a sketch of the prototype magnet used in this work. It is a solenoid plunger magnet having an *E*-shaped structure. A movable plunger seals on one part of the *E* frame and is separated from the outside legs of the *E* frame by a minor air gap.

The field of the magnet is separated into three components which are to be measured from analog and direct models. These fields are indicated in Fig. 2. The first component field results from a current sheet surrounding the plunger. This current sheet is indicated by the heavy lines shown in Fig. 2(A). The height of the current is designated as p, and is seen to be a function of plunger position and consequently a function of the major gap g. The equivalent permeance of the field produced by the current sheet just described is designated as $\mathcal{P}_p$. A current sheet surrounding the major air gap, indicated by the heavy lines in Fig. 2(B), produces the second component field. The height of the current sheet is equal to g, and the equivalent permeance of the field associated with this current sheet is designated as $\mathcal{P}_g$.

The two current sheets described in the foregoing are inductively coupled, giving rise to a third component field, whose permeance is designated as $\mathcal{P}_m$. The mutual permeance $\mathcal{P}_m$, however, is not designated as a major component because it is indirectly determined from other permeances. A detailed discussion of $\mathcal{P}_m$ is deferred to a later part of this paper.

The third major component field results from the magnet coil itself and a balancing current sheet as shown in Figure 2(C). Relative instantaneous polarities of the currents are indicated by the dot-cross symbols on the sketch. The balancing current sheet has a height h, equal to the coil height, and is related to the heights of the current sheets surrounding the major gap and the plunger respectively, by the equation

$$h = g + p \qquad (2)$$

The equivalent permeance of the field resulting from the magnet coil and its balancing current sheet is designated as $\mathcal{P}_w$.

When the component fields described in the foregoing are combined by superposition they leave as their total sum the field of the prototype magnet.

Development of the Analog Model

The fields resulting from the current sheets surrounding the plunger and the major gap are Laplacian fields and can be simulated by electroconducting analog models. Before discussing in detail how a particular model can be constructed to simulate such a field, a brief discussion of the general correspondence between electrical and magnetic phenomenon is in order.

In its simplest form, Ohm's law for electric circuits can be written as

$$G = I/V \qquad (3)$$

where

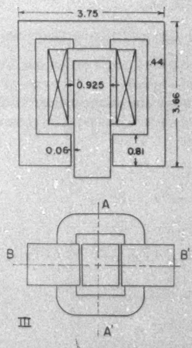


Fig. 1. Prototype magnet showing planes of symmetry

Paper 59-235, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted May 2, 1958; made available for printing December 4, 1958.

J. F. H. Douglas and R. J. Voith are with Marquette University, Milwaukee, Wis.

The authors are greatly indebted to Mr. C. T. Evans of Cutler Hammer, Inc., who initiated this project and assisted greatly in its development.

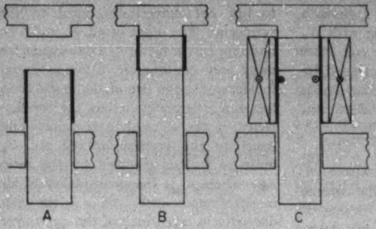


Fig. 2. Fields of magnet

A—Current sheet surrounding plunger
B—Current sheet surrounding major gap
C—The magnet coil and its balancing current sheet

V = potential difference, volts
I = current, amperes
G = conductance of the medium, mhos

Ohm's law for magnetic circuits can be written as

$$\mathcal{P} = \phi/U \tag{4}$$

where

ϕ = flux, webers
U = magnetic potential difference, ampere-turns
$\mathcal{P}$ = permeance of the medium, henrys

Because the equations previously given are identical in form, it is apparent that conductance in an electric circuit is the analog of permeance in a magnetic circuit. It is also apparent that current and potential difference in the electric circuit are respectively the analogs of flux and potential difference in the magnetic circuit. Hence, if a correspondence can be established between ϕ and U in the magnetic circuit and I and V respectively in an equivalent electric circuit, then the electrical quantities may be measured, the conductance computed, and the permeance in the magnetic circuit obtained by analogy.

The manner in which a correspondence between these quantities is effected for the component field resulting from the current sheet surrounding the plunger, will now be described.

Symmetry of the Field

An inspection of Fig. 1 shows that the field of the magnet, as well as all component fields, possess planes of symmetry. The planes designated as AA' and BB' divide each field into four identical parts. Since the equivalent permeance of one of its identical parts differs from the equivalent permeance of the entire field by only a constant multiplier, only one quarter of the field was simulated in constructing the analog models Quadrant III in Fig. 1 shows which quarter of the field was selected in constructing the electroconducting analog models.

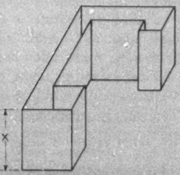


Fig. 3. Model frame

The Conducting Medium

If the field produced by current sheet surrounding the plunger is to be simulated electrically, several things must be accomplished. First, all flux paths in the magnetic field must be replaced by conducting media. Since flux does not wholly confine itself to its intended ferromagnetic circuit, this includes the space surrounding the magnetic structure as well. Secondly, the magnetic potential distribution along the plunger must be known, and a means of duplicating it electrically must be found. And third, a means of accounting for partial flux-linkages in determining the equivalent permeance must be discovered. The second and third of these are accomplished with the aid of a special transformer. The first is accomplished as follows.

Magnetic flux in the magnet exists in two possible media. One medium consists of the ferromagnetic parts of the magnet, that is, its frame and plunger. The other medium consists of the minor and major air gaps and all leakage paths. The permeability of the latter medium is designated as μ_0, and is equal to 3.19 lines/ampere-inch. The permeability of the ferromagnetic medium is by comparison a very large number and is assumed to be infinite. An electroconducting analog model can be effected in this instance by constructing a scale model of the frame out of a material having a very low value of resistivity. The model could then be immersed in an electrolytic bath, thus furnishing the second medium corresponding to the nonferromagnetic space in and around the magnet. If the approximation regarding the permeability of the core in the magnet is to hold in the model, then the resistivity of the model frame must be sufficiently low so that the potential difference between any two points on the frame can be assumed to be zero. In this work the model frame was fabricated from sheet steel as shown in Fig. 3. The model corresponds to that portion of the magnet contained in quadrant III in Fig. 1 since only one quarter of the field is being considered. To facilitate construction and subsequent electric connections, the model was built to a scale k times as large as the magnet. The corresponding reduction in permeance is accounted for by the same factor. The dimension X in Fig. 3 was made larger than required so that the model could be immersed in the electrolyte to its proper depth without submerging the entire model, thus providing easy access to the model during testing.

The Magnetomotive Force Function

The second step in constructing the model is that of simulating the magnetic potential along the current sheet surrounding the plunger. The magnetic potential on the surface of the plunger under the current sheet varies linearly with distance along the plunger as indicated by the straight line graph in Fig. 4. An approximation to this linear potential function is indicated by the step function on the same graph. Obviously, the greater the number of steps, the closer the approximation. Such a potential function was obtained by constructing the model plunger with thick steel laminations, insulated from one another at equal intervals, thus dividing the plunger into sections, and exciting each section from the secondary of a trans-

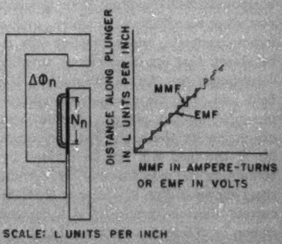

ΔΦn
Nn
MMF
EMF
DISTANCE ALONG PLUNGER IN L UNITS PER INCH
MMF IN AMPERE-TURNS OR EMF IN VOLTS
SCALE: L UNITS PER INCH

Fig. 4. Model showing typical flux linkage, and potential distribution along plunger

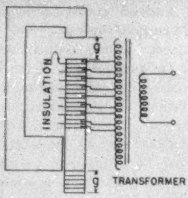


Fig. 5. Transformer connections to segments on model plunger corresponding to a major gap of g inches

former having a sufficient number of equispaced taps as shown in Fig. 5. In order to match the "half steps" at each end of the potential function, each section was made up of two laminations as shown in Fig. 5. The several sections which have been removed to form the major gap have been added to the opposite end of the plunger in order to account for the fact that the plunger actually extends beyond the frame as the gap opens. This artifice is necessary because the fringe permeance of the minor gap is a function of plunger position. The plunger segments were insulated from one another by thin Bakelite laminations which protrude slightly from the plunger surface as indicated in Fig. 5. The amount of protrusion is vital since if the insulation projects too much, the measured conductance will be too small, and vice versa. There are several methods by which the correct projection of the insulation from the plunger surface could be found, such as graphical or experimental mapping or the Schwartz-Cristoffel theorem. An experimental method involving a two-dimensional model cut from conducting paper was used in this work. The procedure is described in Appendix II.

Partial Flux Linkages

The model which is to represent the component field resulting from the current sheet surrounding the model plunger is now complete, and it remains only to be shown that its conductance actually corresponds to the permeance of the field and how the conductance can be measured for any plunger position. An expression for the equivalent permeance $\mathcal{P}_p$ of the field will not be derived by computing the inductance of the current sheet shown in Fig. 4. Assume that the current sheet is comprised of a current ribbon of differential thickness but finite width and is formed by winding the ribbon around the plunger in the manner of a coil. Assume further that the coil so formed has N turns. Fig. 4, shows a small tube of flux $\Delta\phi_n$ linking with N_n of the N turns in the current coil. Let the permeance of the path occupied by $\Delta\phi_n$ be $\Delta\mathcal{P}_n$. Then from equation 1, the inductance ΔL_n resulting from $\Delta\phi_n$ linking with N_n turns is given as

$$\Delta L_n = \Delta\mathcal{P}_n(N_n)^2(10^{-8}) \tag{5}$$

The total inductance L_p, of the current sheet, is then given to any degree of approximation by the following equations

$$L_p = \sum_{n=1}^{m} \Delta L_n = \sum_{n=1}^{m} \Delta\mathcal{P}_n(N_n)^2 10^{-8}$$
$$= \mathcal{P}_n N^2 10^{-8} \tag{6}$$

where $\mathcal{P}_{ps}$ is the total equivalent permeance of the field. Dividing equation 6 by $N^2(10^{-8})$ results in the following expression for $\mathcal{P}_p$

$$\mathcal{P}_p = \sum_{n=1}^{m} \Delta\mathcal{P}_n(N_n/N)^2 \tag{7}$$

Consider now the model which has been constructed to simulate the field produced by the current sheet surrounding the plunger. The ferromagnetic parts of the magnet have been replaced by a virtual short circuit to electric potential. The nonferromagnetic space has been replaced by an electrolyte. The linear magnetic potential along the plunger has been replaced by an approximate step potential which is obtained from the secondary of a transformer having large number of equally spaced taps. Consider the load on the secondary of the transformer. A tube of current Δi_n (analogous to the tube of flux shown in Fig. 4) links with N_n turns of the transformer secondary. Assume that the path occupied by Δi_n has a conductance ΔG_n. The conductance ΔG_n seen from the primary side of the transformer would appear as $\Delta G_n(N_n/N)^2$ provided the number of turns in the primary is N. Since all of the current tubes so defined are in parallel, the total conductance G_p seen from the primary side of the transformer is the sum

$$G_p = \sum_{n=1}^{m} \Delta G_n(N_n/N)^2 \tag{8}$$

A comparison of equations 7 and 8 shows that they are identical in form, and hence the analogy is complete. Therefore, when the electric input to the transformer primary is measured and the equivalent conductance computed, the equivalent permeance of the field is obtained by analogy.

The conductance G_p is proportional to but not numerically equal to the permeance $\mathcal{P}_p$. First of all, G_p must be multiplied by a factor $4/k$ to account for using only one quarter of the field and for increasing the size of the model by a factor k. Secondly, unless the conductivity of the electrolyte is numerically equal to the permeability of air, G_p must be further modified by the ratio μ_0/σ where μ_0 is the permeability of air in lines/ampere-inch, and σ is the conductivity in mhos/inch of the electrolyte used in the model. Hence

$$\mathcal{P}_p = k\mu_0 G_p/4\sigma = k\mu_0\rho\,\frac{G_p}{4} \tag{9}$$

where ρ is the reciprocal of σ or the resistivity of the electrolyte in ohm-inches.

The conductance equation 8 matches the permeance equation 7 exactly. However, the degree to which $k\mu_0\rho G_p/4$ approximates numerically the true equivalent permeance of the field depends on the number of steps used in approximating the potential function on the plunger. In this work it was estimated that the minimum number of steps which resulted in a negligible error was ten.

Circuitry and Measurements

The equivalent permeance $\mathcal{P}_p$, of the component field as obtained from equation 9, is shown in Fig. 6 plotted against the air gap in the prototype magnet. $\mathcal{P}_p$ is of course a function of plunger position. Hence, a series of data must be taken during which the model is modified to simulate different plunger posi-

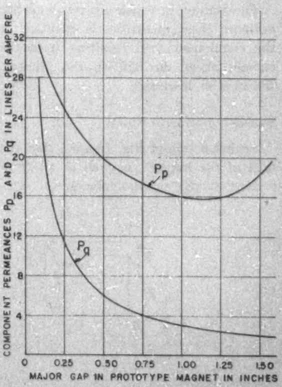


Fig. 6. Permeances $\mathcal{P}_p$ and $\mathcal{P}_q$ of fields produced by current sheets surrounding plunger and major gap respectively

tions, or what is the same thing, different values of major gap. Fig. 5 shows the model adjusted to simulate the plunger position corresponding to a gap of g inches. The gap has been simulated by removing the transformer excitation from the sections which make up the length g, and physically moving the laminations to the opposite end of the plunger. When the laminations are placed on the opposite end of the plunger, the insulation between sections must, of course, be removed. The number of plunger positions which can be simulated in this manner is obviously limited to the number of sections making up the plunger.

The circuitry used in measuring the conductance G_p is shown schematically in Fig. 7. When switch S is placed in position 1, the wattmeter W reads the power into the transformer and the voltmeter V reads the voltage across the primary winding. If for a particular plunger position on the model the wattmeter and voltmeter readings are respectively W_1 and V_1, then the conductance G_{p1} is given as

$$G_{p1}=(W_1-P_1)/V_1^2=W_1'/V_1^2 \qquad (10)$$

where P_1 is the core loss in the transformer at the voltage V_1, and W_1' is the corrected power.

If the voltage or power lost in the transformer windings is not negligible, it will be necessary to further adjust the measured quantities in order to obtain the correct value for G_p. In this work the copper loss was taken into account by substracting from W_1' the quantity $(W_1/V_1')^2(R_1+a'R_2)$ before using it in equation 10. R_1 and R_2 are respectively the total primary and secondary resistance of the transformer windings in ohms, and a' is the variable ratio of the total primary turns to secondary turns used in exciting the model. When the switch S is placed in position 2, V and W read respectively the power and voltage to the device called the "calibrating cell" in Fig. 7. The calibrating cell is a device used to keep a running check on the resistivity of the electrolyte. It consists of a hollow tube of insulating material with a conducting plate at each end of the tube. The tube contains several small holes which allow it to become filled when immersed in the electrolyte and to drain when removed. If the inside dimensions of the cell are known, it contains a known volume of electrolyte when immersed, and if the conductance of this volume can be measured the resistivity of the electrolyte can be computed from the equation

$$\rho=A_e/G_e l \qquad (11)$$

where

ρ=resistivity of the electrolyte, ohm-inches
G_e=conductance, mhos of the electrolyte contained in the cell
l and A_e=inside dimensions of the tube, inches and square inches respectively

The conductance of the cell is obtained by dividing the square of the voltage drop across the cell into the power consumed by the cell. Hence, in this manner the resistivity of the electrolyte was checked immediately after each reading was taken on the model, thus avoiding errors due to variations resulting from changes in temperature and from settling. The electrolyte used in these experiments was a solution of sodium carbonate (Na_2CO_3) concentrated to approximately one part normal solution in 60 parts of water.

Component Permeance $\mathcal{P}_q$

The permeance of the component field resulting from the current sheet surrounding the major gap shown in Fig. 2(B) is obtained partly by calculation and partly by experimentation. If it is again assumed that no ampere-turns are consumed in the iron parts of the magnet, the equivalent reluctance of this field consists of the reluctance of the major gap in series with the reluctance of the minor gap and its fringe area. The field inside the current sheet is regular and uniform and its reluctance $\mathcal{R}_g$ can be computed from the formula

$$\mathcal{R}_g=g/\mu_0 A \qquad (12)$$

where g and A are the gap dimensions of the major gap in inches and square inches respectively, and μ_0 is the permeability of air in lines/ampere-inch. The reluctance $\mathcal{R}_f$ of the minor gap and its fringe is not a fixed quantity but is a function of plunger position. As the plunger moves out from its position at zero gap the proximity of iron at the minor gap changes thus altering the field and hence its reluctance.

The reluctance $\mathcal{R}_f$, or its reciprocal $\mathcal{P}_f$, can be obtained in a manner similar to the way $\mathcal{P}_p$ was obtained, using the same model with one modification. It will be observed that in the model used to determine $\mathcal{P}_p$, the permeance at zero gap is equal to $\mathcal{P}_f$. This is true since all sections making up the plunger are in place and under excitation from the transformer secondary, and hence, no major gap exists in the model and the entire electromotive force is consumed in the minor gap and its fringe area.

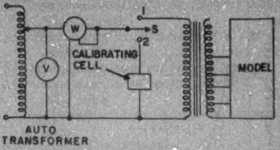


Fig. 7. Circuit used for testing analog models

The variation in $\mathcal{R}_f$ with gap length may be simulated by adding laminations to the model plunger, thus extending the plunger beyond its position at zero gap without actually creating a major gap by removing laminations from the plunger proper. Thus, the reluctance obtained from electrical quantities measured on the primary side of the transformer in the usual manner is the reluctance of the minor gap and its changing fringe only.

The accuracy of the procedure described with the foregoing for obtaining $\mathcal{R}_f$ may be questioned since the presence of a major gap may in fact change the reluctance. However, since $\mathcal{R}_f$ is a very small percentage of the total reluctance, an error can be tolerated in $\mathcal{R}_f$ without introducing an appreciable error in the total reluctance, and therefore the procedure was not checked as a major source of error. The total equivalent reluctance $\mathcal{R}_q$ of the component field is the sum of $\mathcal{R}_f$ and $\mathcal{R}_g$, and the equivalent permeance $\mathcal{P}_q$ of this field is therefore the reciprocal of $\mathcal{R}_q$. In terms of the reluctances given previously

$$\mathcal{P}_q=1/\mathcal{R}_q=1/(\mathcal{R}_g+\mathcal{R}_f) \qquad (13)$$

The permeance $\mathcal{P}_q$ is shown as a function of the air gap in the prototype magnet in Fig. 6.

Mutual Permeance

As stated earlier the two current sheets surrounding the plunger and the major gap give rise to component fields which are mutually coupled, and therefore their permeances as just obtained must be adjusted to account for this fact. It was also stated previously that the two current sheets were electrically connected in series with their fields reinforcing one another. The equivalent inductance L_t of such a system is given as

$$L_t=L_q+L_p+2M \qquad (14)$$

where the subscripts q and p have their previous significance and M is the mutual inductance in henrys. The individual inductances in equation 14 are functions of field permeance and may be replaced by their respective permeances provided

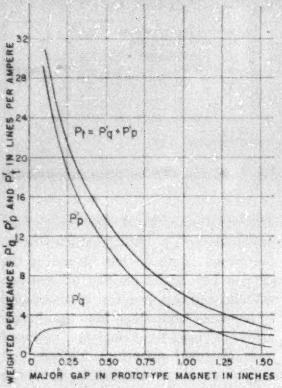


Fig. 8. Permeances $\mathcal{P}_p$ and $\mathcal{P}_p$ modified to account for mutual inductance between current sheets

a suitable weighting factor corresponding to N^2 in equation 1 can be found.

Assume for the moment that the current sheets surrounding the major gap and the plunger have a finite number of turns given as N_q and N_p respectively and that the total turns of the two coils thus formed is N_t. In accordance with elementary circuit theory, the terms in the right hand member of equation 14 may be transformed as follows

$$L_t = N_q^2\mathcal{P}_q + N_p^2\mathcal{P}_p + 2N_qN_p\mathcal{P}_m \qquad (15)$$

Now if equation 15 is divided by N_t^2, the left-hand member becomes the total equivalent permeance $\mathcal{P}_t$ of the two fields acting together, and the terms in the right-hand member become their respective field permeances modified by an appropriate turns-ratio. That is

$$\mathcal{P}_t = (N_q/N_t)^2\mathcal{P}_q + (N_p/N_t)^2\mathcal{P}_p + 2N_qN_p\mathcal{P}_m/N_t^2 \qquad (16)$$

Since the turns-rations in equation 16 in reality represent ratios of linear current distribution, they may be replaced by ratios of their respective heights. Making use of the relationship and symbols defined in equation 2, equation 16 can be written

$$\mathcal{P}_t = (g/h)^2\mathcal{P}_q + (p/h)^2\mathcal{P}_p + 2qp\mathcal{P}_m/h^2 \qquad (17)$$

In determining the mutual inductance of the system previously described, one might reasonably conclude that practically all of the flux produced by the current sheet surrounding the major gap links with current sheet surrounding the plunger since the latter is terminated on either end by iron. If it can be assumed that the current sheet surrounding the major gap produces no leakage flux, it follows that $\mathcal{P}_m = \mathcal{P}_q$ in equation 17. Making this substitution reduces equation 17 to

$$\mathcal{P}_t = p_q' + \mathcal{P}_p' \qquad (18)$$

where

$$\mathcal{P}_q' = (g/h)(2 - g/h)\mathcal{P}_q \qquad (19)$$

and

$$\mathcal{P}_p' = (p/h)^2\mathcal{P}_p \qquad (20)$$

The three permeances $\mathcal{P}_q$, $\mathcal{P}_p$, and $\mathcal{P}_t$ are shown in Fig. 8 as a function of the major air gap in the prototype magnet.

Component Permeance $\mathcal{P}_w$

The remaining permeance to be considered is that of the field resulting from the magnet coil and its balancing current sheet shown in Fig. 2(C). Because of the presence of current, the resulting magnetic field is nonLaplacian. Hence, an electroconducting analog model cannot be used since no potential function can be specified. In this case the permeance is determined from electrical measurements made on a suitable model which generates the field directly. The model used therefore, is not an analog model, but merely a test model using iron and copper.

The test model was formed by constructing a special 3-winding coil having N_w turns in each winding. Winding *1* was placed on a Bakelite block having dimensions proportional to the prototype plunger. This winding was closely wound with fine wire no. 26 Awg (American Wire Gage) so as to approximate a current sheet on the plunger surface. A Bakelite frame having outside dimensions proportional to the inside dimensions of the prototype coil was then placed around winding *1*. Two strands of no. 14 Awg wire were then wound together on the outer Bakelite frame forming two curls. Each curl had the same number of turns as winding *1*.

The active winding (winding *2*) corresponds to the prototype coil, and the remaining winding (winding *3*) was utilized as a test winding. For convenience the model was made k_w times as large as the prototype. The 3-winding coil described in the foregoing was then placed inside a laminated iron frame having dimensions k_w times those of the prototype but with no plunger. Winding *1* and winding *2* were then connected in series so as to oppose one another and the series combination energized from a 60-cycle source. The resulting voltage induced in winding *3* was measured with a vacum tube voltmeter. The voltage induced in winding *3* results from leakage flux between the thin coil *1* and the thick coil *2*. Hence the ratio of this voltage to the angular velocity times the current in windings *1* and *2*, is the leakage inductance L_w or the equivalent inductance of a thick coil and its balancing current sheet. The equivalent permeance $\mathcal{P}_w$ is then given as

$$\mathcal{P}_w = L_w(10^8/N_w^2k_w) \qquad (21)$$

The permeance $\mathcal{P}_w$, obtained by the method described was 2.8 lines/ampere for the prototype.

Since the coil and balancing current sheet produces no field inside the current sheet, the permeance obtained in the manner described would be the same regardless of what was inside the current sheet, hence, the plunger was omitted in constructing the test model. However, in the prototype model the partial flux (which the test model generates) is not restricted to paths parallel to the plunger, and in fact may be shorter in length when the plunger is present inside the coil. The permeance $\mathcal{P}_w$ can therefore be expected to be a function of plunger position, decreasing slightly with increasing air gap to a minimum of 2.8 lines/ampere at maximum gap. Since $\mathcal{P}_w$ was assumed constant in this work, the minimum value of 2.8 lines/ampere was selected so as to introduce the least error in the total permeance. In other words, permitting the largest error in $\mathcal{P}_w$

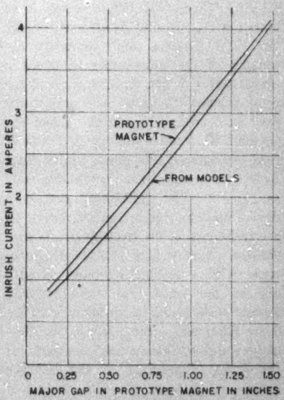


Fig. 9. Inrush current as determined from tests on models and by measurement on prototype magnet

to occur where $\mathcal{P}_w$ is the smallest percentage of the total permeance. A reasonably close approximation to $\mathcal{P}_w$ can easily be obtained mathematically with the aid of several simplifying assumptions as shown in the Appendix I.

The total equivalent permeance of the prototype magnet is given as

$$\mathcal{P}_{eq}=\mathcal{P}_t+\mathcal{P}_w \tag{22}$$

and is identical in shape to the curve $\mathcal{P}_t$ shown in Fig. 8 since $\mathcal{P}_w$ is a constant.

Comparison of Results with Prototype

In order to afford a simple means of comparison, the inrush current of a magnet whose field has the permeance given in equation 22 is computed from the following equation

$$I_r=E(R^2+X_{eq}^2)^{-1/2} \tag{23}$$

in which X_{eq} has the value

$$X_{eq}=2\pi fL_{eq}=2\pi N_w^2(10^{-8})\mathcal{P}_{eq} \tag{24}$$

where

E=rated voltage of the prototype magnet
N_w, R=number of turns and the resistance in ohms, respectively, of its coil
f=frequency of the applied voltage
I_r=inrush current, amperes

The inrush current calculated in the manner described is shown in Fig. 9 along with the inrush current actually measured on the prototype magnet. The discrepancy between the two curves at any value of gap is less than 10% of the inrush current. Some discrepancy is to be expected since the ampere-turns consumed in the iron were neglected in computing the current. It is estimated that neglecting the ampere-turns consumed in the iron results in an error of approximately 0.1 ampere. This cuts the discrepancy in Fig. 9 to about one half.

Conclusion

A fair degree of agreement between tests on the models and the prototype is to be noted. The discrepancies indicate the need for further study and experiment. The results indicate also that the analysis of other three-dimensional field problems using the artifice of current sheets might be worthy of investigation.

Nomenclature

A=area of major gap, inches^2
A_c=area of calibrating cell, inches^2
E=rated voltage of prototype coil, volts
G=conductance, mhos
G_c=conductance of calibrating cell, mhos
G_p, G_{pi}=equivalent conductance of field due to plunger current sheet, mhos
I=current, amperes
I_r=inrush current of prototype magnet, amperes
L=inductance, henrys
L_{eq}=equivalent inductance of prototype magnet, henrys
L_p, L_q=equivalent inductance of plunger and major gap current sheets respectively, henrys
L_t=equivalent inductance of mutually coupled current sheets, henrys
L_w=equivalent inductance of a thick coil and its balancing current sheet, henrys
M=mutual inductance, henrys
N=turns of a coil, dimensionless
N_p, N_g, N_t=fictitious turns of plunger and major gap current sheets, and their sum respectively, dimensionless
N_w=turns of prototype coil, 984 turns
P_1=transformer core loss at V_1 volts, watts
$\mathcal{P}$=permeance, henrys
$\mathcal{P}_{eq}$=equivalent permeance of prototype magnet, lines/ampere
$\mathcal{P}_f$=permeance of minor gap (lines/ampere)
$\mathcal{P}_m$=equivalent permeance of field proposed to account for inductive coupling between current sheets, lines/ampere
$\mathcal{P}_p$, $\mathcal{P}_g$=equivalent permeance of field due to plunger and major gap current sheets respectively, lines/ampere
$\mathcal{P}_p'$, $\mathcal{P}_g'$,=$\mathcal{P}_p$ and $\mathcal{P}_g$ respectively, modified to account for mutual coupling between current sheets, lines/ampere
$\mathcal{P}_t$=equivalent permeance of total current sheet, lines/ampere
$\mathcal{P}_w$=equivalent permeance of field due to thick coil and a balancing current sheet, lines/ampere
R=resistance of prototype coil, ohms
R_1, R_2=resistance of transformer primary and secondary windings respectively, ohms
$\mathcal{R}_g$=reluctance of major air gap, amperes/-line
$\mathcal{R}_f$, $\mathcal{R}_g$=reciprocals of $\mathcal{P}_f$ and $\mathcal{P}_g$ respectively, amperes/line
U=Magnetic potential difference, amperes
V=electric potential difference, volts
V_1=meter reading, volts
W_1, W_1'=meter reading and corrected meter reading respectively, watts
X_{eq}=reactance of prototype magnet, ohms
a'=transformer turns ratio, dimensionless
f=frequency, 60 cycles/second
σ=length of major gap and its current sheet, inches
h=height of total current sheet, inches
k, k_w=scale factors, 2.7 and 2.0, respectively
l=length of calibrating cell, inches
p=height of current sheet surrounding plunger, inches
$\Delta\phi_n$=elemental flux tube, lines
$\Delta\mathcal{P}_n$=permeance of path occupied by $\Delta\phi_n$, lines/ampere
ΔL_n=inductance associated with $\Delta\mathcal{P}_n$, henrys
Δi_n=elemental tube of current, amperes
ΔG_n=conductance of path occupied by Δi_n, mhos
μ_0=permeability of air, 3.19 lines/ampere-inch
ϕ=magnetic flux, webers
ρ=resistivity of electrolyte, ohm-inches
σ=conductivity of electrolyte, mhos/inch

Appendix I

Computed Value of Permeance $\mathcal{P}_w$

Fig. 10 shows a sketch of the magnet coil and its balancing current sheet along with the assumed flux density distribution in a plane passing through the axis of the coil. The flux is assumed to exist in straight lines parallel to the axis of the coil and having a length H-δ, where δ must be chosen on the basis of experience or experiment. The permeance $\mathcal{P}_{wc}$ of that portion of the field between the coil and the current sheet can be written down immediately as

$$\mathcal{P}_{wc}=\mu_0(cd+ab)/(H-\delta) \tag{25}$$

since in this region the flux density is uniform and the flux links completely all of the turns of the coil. The permeance of the differential tube of width dx shown in Fig. 10(A) is given as

$$d\mathcal{P}_{wp}=\mu_0(x/T)^2(1-2\pi x)dx/(H-\delta) \tag{26}$$

where x/T gives the ratio of turns linked by the tube to the total turns in the coil, and l is the length of the outside perimeter of the coil and is equal to

$$1=2(c+d)+2\pi T \tag{27}$$

The permeance of that portion of the field inside the coil is therefore this integral of $d\mathcal{P}_{wp}$ between the limits of $x=0$ and $x=T$.

$$\begin{aligned}\mathcal{P}_{wp}&=\int_0^T\mu_0(x/T)^2(1-2\pi x)dx/(H-\delta)\\&=\mu_0T(1/3-2\pi T/4)/(H-\delta)\end{aligned} \tag{28}$$

The total permeance is then the sum of $\mathcal{P}_{wc}$ and $\mathcal{P}_{wp}$ given in the foregoing that is

$$\mathcal{P}_w=\mathcal{P}_{wc}+\mathcal{P}_{wp} \tag{29}$$

Taking the dimensions given in Fig. 1, and $\delta=0$, $\mathcal{P}_w$ has the value 2.69 lines/ampere. The value of 2.8 lines/ampere obtained experimentally for $\mathcal{P}_w$ indicates that δ is 3.89% of H.

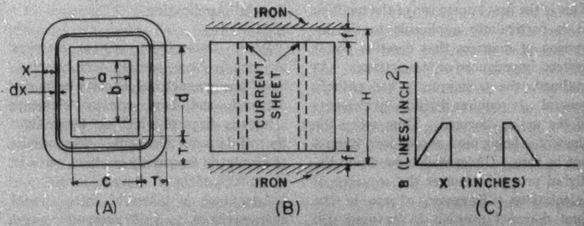


Fig. 10. Flux density in thick coil with balancing current sheet

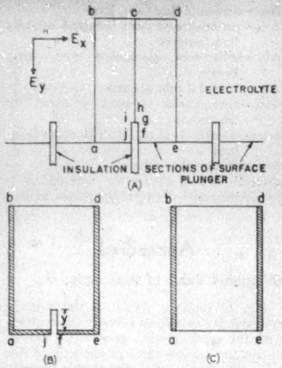


Fig. 11. Field models representing component field E_x for

A—Field alone showing three insulating laminations
B—Step potential
C—Uniformly varying potential

Appendix II

The magnetic potential varies uniformly along the surface of the plunger in the prototype magnet, but the analogous electric potential in the analog model varies in finite steps as shown in Fig. 4. The sections of the model plunger are insulated from one another by thin laminations of Bakelite which protrude slightly from the surface of the iron. If the length which these laminations protrude from the surface is selected properly, the resistance of the field resulting from the step potential can be made equal to the resistance of the field due to a uniformly varying potential. In the discussion which follows it is assumed that the proper length can be found on the basis of a two-dimensional analysis.

A portion of the model plunger showing three insulating laminations which are exaggerated in thickness is shown in a plane perpendicular to the plunger surface in Fig. 11(A). The component of the electric field intensity E_y perpendicular to the plunger surface is not affected by the length y of the protruding insulation. The protrusion y does however affect the field E_x parallel to the plunger surface. If the field E_x is considered alone, a plane perpendicular to the plunger surface contains the equipotential lines *jab*, *hc*, *fed*, and the flow lines *bcd* and *ajigfe* as shown in Fig. 11(A). A two dimensional model of a field having boundaries described by the equipotential and flow lines previously cited can be made out of conducting paper as shown in Fig. 11(B) where the edges *jab* and *fed* have been painted with silver conducting paint. The resistance R_f between the equipotential lines of the model can be determined as a function of the protrusion y. The resistance R_u of a field due to a uniformly varying potential rather than a step potential can be obtained from the model shown in Fig. 11(C) in which *ab* and *cd* are equipotential lines and *bd* and *ad* are flow lines. If the resistance R_u is equal to the resistance R_f, there is no difference between the step potential on the model and the linear potential on the prototype plunger as far as the equivalent resistance or reluctance is concerned.

Since the resistance R_f is a function of the cut y (which corresponds to the amount of protrusion of the insulation into the electrolyte), R_f can be made equal to R_u by adjusting y to its proper length by a trial and error process. It was found experimentally by the foregoing method that the amount which the insulation should protrude into the electrolyte was 56% of the dimension *aj*.

◆

Discussion

Thomas J. Higgins (University of Wisconsin, Madison, Wis.): This paper is most interesting. It provides yet another striking illustration of the value of electrolytic tank and of conducting paper analog procedures in gaining solution of problems of current interest in industry, which are rather intractable to quick solution by analytic procedures.

In particular, the use of current sheets for separating non-Laplacian from Laplacian components, and the use of boundaries with step function potential to replace a linear variation of potential, may afford convenient analog solutions in many cases.

J. F. H. Douglas and **R. J. Voith:** While paper models can be used to solve some field problems in and outside copper windings, we were unable to find any conducting model to measure our third component of field. We conclude that the method we used will have to suffice until such conducting models are found.

Although the inductance of a magnet is related to its force-stroke curve, we did not discuss this topic in the paper because there seems to be no one generally accepted method of measurement. The three experimental methods and the calculated values from our tests differed materially from each other.

We wish to say in conclusion that force-stroke curves are important. Standards of measurement should be set up and further analytic studies should be made so that practical designs can be made.

Vol.
Pt. 1
195

A New Direct Matrix Inversion Method

R. BRUCE SHIPLEY
FELLOW AIEE

DOROTHY COLEMAN
NONMEMBER AIEE

THIS PAPER describes a special-purpose matrix inversion technique developed in the System-Planning Branch of TVA (Tennessee Valley Authority). This is the first known use of the method. It is particularly applicable to the inversion of matrices that describe power system impedances or admittances. As outlined, the technique is not entirely general. It requires division by elements of the major diagonal. Thus, when the element is being used as a divisor, it must not be zero. Under such a condition the digital program changes the sequence of calculations. However, a zero in the final diagonal position to be used will cause the matrix to be considered singular. The technique can be made more general but this is not necessary for most power system work.

Special Application

Power transmission system matrices are usually large, contain complex numbers, are symmetrical about the major diagonal, and have nonzero elements along the diagonal which are larger than the off-diagonal elements. The admittance matrix frequently contains zeros in about 90% of the off-diagonal terms.

Advantage is taken of these special characteristics to gain computer speed and to reduce the high-speed memory requirements. This is partially accomplished by storing only the diagonal and one half of the nondiagonal elements. Only one half of the nondiagonal elements are calculated and zero recognition is used to improve speed. The method is believed to be faster than other methods available to the industry.

Inversion of such matrices is desirable in economic studies, short-circuit studies, equivalent circuit studies, etc.

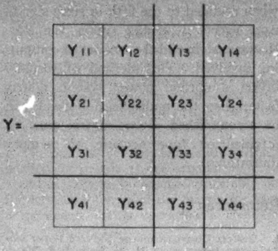
The Method

To clarify the explanation an entire matrix will be used rather than only one

---

Paper 59-502, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Southeast and South Central District Meeting, Atlanta, Ga., April 8-10, 1959. Manuscript submitted December 18, 1958; made available for printing February 13, 1959.

R. Bruce Shipley and Dorothy Coleman are with the Tennessee Valley Authority, Chattanooga, Tenn.



$$Y=\left[\begin{array}{cc|c|c} Y_{11} & Y_{12} & Y_{13} & Y_{14} \\ Y_{21} & Y_{22} & Y_{23} & Y_{24} \\ \hline Y_{31} & Y_{32} & Y_{33} & Y_{34} \\ \hline Y_{41} & Y_{42} & Y_{43} & Y_{44} \end{array}\right]$$

Fig. 1. Matrix to be inverted

half of the matrix. The inversion rules are very simple and will be explained in detail for only one diagonal element. Let the diagonal element Y_{33} of Fig. 1 be selected for explanation. The rules are as follows:

1. All elements external to the row and column containing the diagonal element must be modified. These elements are modified one at a time.

For example, Y_{11} is modified to become

$$Y_{11}-\frac{Y_{13}Y_{31}}{Y_{33}} \quad (1)$$

The element Y_{14} is modified to become

$$Y_{14}-\frac{Y_{13}Y_{34}}{Y_{33}} \quad (2)$$

The element Y_{42} is modified to become

$$Y_{42}-\frac{Y_{32}Y_{43}}{Y_{33}} \quad (3)$$

The element Y_{44} is modified to become

$$Y_{44}-\frac{Y_{34}Y_{43}}{Y_{33}} \quad (4)$$

The other external elements are similarly modified.

2. The element Y_{33} is replaced by $-1/Y_{33}$. Each other element in the row and column common to Y_{33} is multiplied by $-1/Y_{33}$.

3. Rules 1 and 2 must be repeated for each main diagonal element. The resulting matrix will be minus the inverse of the original matrix. The operations may be carried out in any sequence. Note that symmetry is not a requirement when the entire matrix is used.

Adaptability to Computer

The method described was originally devised to invert matrices in the high-speed memory of a 16,000-word IBM (International Business Machine Corporation) *704* computer. Symmetrical matrices in complex numbers slightly larger than 100 by 100 can be inverted by this computer. In addition to the program and the matrix in triangular form, it is necessary to provide additional memory location for only one matrix element. For convenience in checking the inverse, space is provided for one row of the original matrix. Thus, the memory requirements and the numerical work required has been minimized.

The method is adaptable to the inversion of larger matrices by the use of tape. In this case, it is probably best to store each complete row of the matrix as a record on a binary tape. Thus it becomes possible to perform arithmetic operations on one row of the matrix at a time. Such a procedure limits the size of the matrix only by time requirements and by round-off error when this becomes a problem.

For very large matrices the method is adaptable to partitioning. The matrix may be partitioned so that square submatrices are formed along the main diagonal. The submatrices need not all be the same size. The submatrices along the diagonal may then be inverted and the complete inversion of the large matrix accomplished by carrying out matrix operations similar to those described for the individual elements.

Derivations

The method described was derived more from a knowledge of electric circuits than from a knowledge of mathematics. A normal job for the a-c network analyzer is to derive an equivalent circuit for a power system holding a smaller number of busses than actually exist on the power system. This is usually done by plugging the power system on the analyzer and connecting each bus to be held through a metering jack to a common bus. Thus, by opening one metering jack at a time and applying a voltage, the equivalent circuit may be determined from the currents that flow in the metering jacks. This procedure is the equivalent of Kron's reduction formula as described by the matrix equation:

$$Z_1-Z_2Z_4^{-1}Z_3 \quad (5)$$

In equation 5, Z_1, Z_2, Z_3, and Z_4 are submatrices resulting from partitioning a large matrix, where Z_1 is the upper left corner, Z_2 the upper right corner, Z_3 the lower left corner, and Z_4 the lower right corner. Z_1 and Z_4 must be square.

If in equation 5, Z_1 and Z_4 are made the same dimension and if $Z_1=0$, and if $Z_2=Z_3=I$ (the identity matrix), then the resulting expression becomes $-Z_4^{-1}$. Now, the same results as expressed in equation 5 can be obtained by eliminating one row and one column at a time. Thus, with the described restraints on Z_1, Z_2, Z_3, and Z_4, it is possible to place the newly calculated elements back in the Z_4 positions. Therefore, the memory space that would normally be required for Z_1, Z_2, and Z_3 is not needed. The simple rules just described are the result of observing the effect of this operation.

In partitioning a very large matrix to be inverted, observe that it may be possible to do so, in such a way that the off-diagonal submatrices contain only zeros. This, of course, will save considerable numerical work. In fact, it supplies some insight into Kron's method of diakoptics. For example, if the matrix to be inverted is an admittance matrix of a transmission system, the off-diagonal elements represent physical lines on the power system. Thus, if certain lines are simply omitted from the power system, the partitioned matrix reduces to compound diagonal form. Therefore, the compound diagonal elements are simply inverted to obtain the matrix inverse less the lines that were omitted. The omitted lines can be added by inverting a matrix whose dimensions are equal to the lines omitted. Subsequent matrix operations (not involving inversion) give the required inverse. Obviously, if too many lines must be left out of the original matrix the process of reinserting them may involve more work than the inversion of the original matrix. Usually diakoptics saves work when there are already many zeros in the system to be inverted.

Possible Extension of the Method

From the description it is apparent that the method cannot be used in its present form to invert a matrix that has all zeros along the main diagonal. Combinatorial topology supplies a possible means of circumventing this difficulty. For example, an arbitrary number is simply inserted for one of the zeros in the diagonal position. The matrix is then inverted and the inverse is modified to remove the arbitrarily inserted number. In mathematical operation, the modification is identical to switching a line in the power system impedance matrix. This procedure has been outlined in reference 1. The matrix modification (or line switching) requires 12 seconds on a 100-by-100 matrix. In electrical engineering terms, the operation is similar to placing a line in service while other work is done on the power system, and then removing the line from service.

Of course, it is possible that changing only one of the zero diagonal elements will not permit the method to invert the matrix. In this case it may be necessary to change arbitrarily more than one of the elements. Of course, it will be necessary to make as many modifications of the inverse as there were lines added to the original system.

This description does not intend to imply that extensive investigations have been carried out to avoid the difficulties that may arise in special cases.

Conclusions

The method described was devised for use with a 16,000-word IBM *704* computer using single precision arithmetic. The machine time required is a function of the number of zeros in the matrix. A 100-by-100 complex matrix inversion with no zeros required 24 minutes of computing time. A normal power system admittance matrix, 99 by 99, required 11 minutes to invert. The technique described has not been extensively used as an inversion method. Its equivalent as a part of several other programs developed by TVA has been used successfully to invert hundreds of matrices of varying sizes up to and including the maximum size that can be accommodated in high-speed memory. Present practice requires that each inverse be checked by multiplying the inverse by the original matrix. Usually this results in an identity matrix when rounded to the fifth or sixth significant figure, depending upon the particular matrix and the amount of computing involved. This method will usually invert a normal power system matrix faster than the multiplication check can be accomplished, if one matrix must be retained on tape because of memory limitations.

References

1. Digital Computation of Driving Point and Transfer Impedances, H. W. Hale, J. B. Ward. *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 76, Aug. 1957, pp. 476–81.

2. Tensor Analysis of Networks (book), Gabriel Kron. John Wiley & Sons, Inc. New York, N. Y., 1939, chap. X.

◆

Discussion

Gordon K. Carter (General Electric Company, Schenectady, N. Y.): In this paper the authors have apparently attained the ultimate in minimum storage for direct inversion of a matrix, along with the maximum speed for such a process. The following comments are intended to be helpful by relating the procedure to the process of elimination of variables, which may be more familiar to some readers than Kron's reduction formula and the associated circuit manipulations. The elimination method has been illustrated[1] by a simple 3-variable numerical example which could be written more formally as follows:

$$\begin{aligned} A_{11}X_1+A_{12}X_2+A_{13}X_3&=Y_1+0+0\\ A_{21}X_1+A_{22}X_2+A_{23}X_3&=0+Y_2+0\\ A_{31}X_1+A_{32}X_2+A_{33}X_3&=0+0+Y_3 \end{aligned} \qquad (6)$$

The inversion process interchanges the forms of the two sides of these equations, so that (X_1,X_2,X_3) can be evaluated by simple multiplications and additions for given (Y_1,Y_2,Y_3), just as the reverse is true for these original equations. This can be done by eliminating X_1, say, from all but the first equation; then X_2, say, from all but the second; etc. Only one such elimination need be described, since the rest follow the same pattern of operations. For instance, to eliminate X_1 from all but the first equation:

1. Divide the first equation by A_{11}.

2. Multiply the resulting expression by A_{21} and subtract from the second equation; multiply by A_{31} and subtract from the third; etc.

3. At this point the expressions appear as follows:

$$\begin{aligned} X_1+&\frac{A_{12}}{A_{11}}X_2+\frac{A_{13}}{A_{11}}X_3=\frac{1}{A_{11}}Y_1+0+0\\ 0\ +&\left(A_{22}-A_{21}\frac{A_{12}}{A_{11}}\right)X_2+\left(A_{23}-A_{21}\frac{A_{13}}{A_{11}}\right)X_3=-\frac{A_{21}}{A_{11}}Y_1+Y_2+0\\ 0\ +&\left(A_{32}-A_{31}\frac{A_{12}}{A_{11}}\right)X_2+\left(A_{33}-A_{31}\frac{A_{13}}{A_{11}}\right)X_3=-\frac{A_{31}}{A_{11}}Y_1+0+Y_3 \end{aligned} \qquad (7)$$

In terms of digital computer operation, storage need be provided only for coefficients which are of an unknown, or variable, character from problem to problem. In expressions 1 and 2 there are no more of these than the number in the original matrix of A_{ij}. Consequently, if in expression 2 the coefficients of Y_1 are stored where the A_{i1} originally were, the storage requirements are at a minimum; but of course a record must be kept to ensure that the elimination ultimately takes place once and only once for each of the original matrix columns.

An embodiment of this procedure as a program for the IBM *650* computer recently came to our attention.[2] That particular program uses extra storage equivalent to one extra row of the matrix; it probably uses fewer operating instructions, but is less flexible, than the authors' method.

Comparison of the elimination procedure with that of Shipley and Coleman shows that the operations are nearly identical. The only material difference is in the appearance of negative signs in both the row and column of the pivotal term in the paper, whereas, in the straight elimination process, terms in the pivotal row carry a positive sign. In the end, this results in the development of the negative of the inverse matrix in the first case, the actual inverse in the second.

As the authors point out, the elimination can proceed in any order. For instance, at each stage the pivotal element can be chosen as the largest of those on the main diagonal which remain to be processed. From the more general point of view, however, it is not necessary to restrict the pivotal elements to the main diagonal. Any set could be used, so long as no two are chosen from the same row or the same column. From the standpoint of error control, for example, it might be desirable to choose for the pivotal element at any step the largest term in any part of the matrix not in conflict with the rule against duplicating rows and columns; the penalty would lie in the need for appropriately rearranging the rows and columns of the final matrix so that it is the proper inverse of the original one.

REFERENCES

1. Some Practical Mathematical Techniques, Gordon K. Carter. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, July 1958, pp. 295–302.

2. Minimum-Storage Matrix Inversion, Victor W. Bolie. *Zietschrift fuer Angewandte Mathematik und Mechanik*, Berlin, Germany, vol. 38, Sept.-Oct. 1958, pp. 369–72.

V. Converti (Arizona Public Service Company, Phoenix, Ariz.): The authors' intriguing new method of matrix inversion is adequate for the inversion of any matrix, in general, and welcome for the inversion of large power system matrices, in particular. The technique developed in the paper is of particular value because it is based on a direct method of inversion, thus avoiding inversion by iteration processes; and because the storage requirement for the input is reduced to the triangular form of the matrix. It therefore represents a significant contribution to the continued effort of increasing the speed of inversion over available general inversion methods, and of reducing input requirements of large power system matrices handled by digital computers.

The authors' method can be applied to the calculation of network equivalents by a modification of the general case. The mathematical operation for the general case is based on the application of rules 1 and 2 of the method to each element of the matrix for successive rows and columns. As outlined by the authors, rule 1 makes use of Kron's reduction formula, which operates on those elements external to the row and column selected, while rule 2 operates on the row and column selected.

Rule 1 is exactly what is needed when one row and one column are eliminated from the matrix. Therefore, restricting the operation to rule 1 only, one row and one column are eliminated each time the reduction formula is applied to all elements affected. By repeating the process on the reduced matrix by one, the next row and column can be eliminated, and so on. This operation corresponds to the familiar network reduction steps performed by hand.

Thus, to obtain driving point and transfer admittances representing an equivalent network for use in a stability study, only two steps are needed with the use of rule 1:

1. Formation of the self- and mutual-admittance matrix of the original network with the nodes to be retained preferably on the left-hand corner of the matrix for ease of reference. (This operation is common to any method.)

2. Application of rule 1 to eliminate all rows and columns containing those nodes to be eliminated.

In addition to the advantages indicated by the authors, of faster inversion and reduced storage requirement, the use of this modified method to calculate network equivalents makes available for the first time a step-by-step record of the network reduction process. This, in effect, would require the writing on the output, if so desired, of the reduced matrix after each node elimination representing the network at that particular stage of computation. The reduced matrix can be made available on tape for further processing, or as a starting point for changes in the network configuration.

H. W. Hale (Wayne State University, Detroit, Mich.) and **R. W. Goodrich** (University of New Hampshire, Durham, N. H.): That the general method has other useful applications may be shown by a description of the program we have developed for a different purpose, but which is similar to the authors's useful technique.

In a recent paper[1] we described a procedure for the digital solution of power flow problems that makes use of network equations of the form

$$\left\| \begin{matrix} I_1 \\ \vdots \\ I_J \\ \\ E_K \\ \vdots \\ E_N \end{matrix} \right\| = \left\| \begin{matrix} Y_{11} \ldots Y_{1J} & D_{1K} \ldots D_{1N} \\ \vdots \qquad \vdots & \vdots \qquad \vdots \\ Y_{J1} \ldots Y_{JJ} & D_{JK} \ldots D_{JN} \\ & \\ D_{K1} \ldots D_{KJ} & Z_{KK} \ldots Z_{KN} \\ \vdots \qquad \vdots & \vdots \qquad \vdots \\ D_{N1} \ldots D_{NJ} & Z_{NK} \ldots Z_{NN} \end{matrix} \right\| \left\| \begin{matrix} E_1 \\ \vdots \\ E_J \\ \\ I_K \\ \vdots \\ I_N \end{matrix} \right\| \tag{8}$$

where the Y's, D's, and Z's are admittances, dimensionless transfer ratios, and impedances respectively. An additional note is made that D_{KJ} is the negative of D_{JK}.

The determination of the elements of the coefficient matrix of equation 8 from either the impedance or admittance matrix is the purpose for which our program was developed. The basic ideas are illustrated by starting with the network equations in the form

$$\left\| \begin{matrix} I_1 \\ \vdots \\ I_K \\ \vdots \\ I_N \end{matrix} \right\| = \left\| \begin{matrix} Y_{11} \ldots Y_{1K} \ldots Y_{1N} \\ \vdots \qquad \vdots \qquad \vdots \\ Y_{K1} \ldots Y_{KK} \ldots Y_{KN} \\ \vdots \qquad \vdots \qquad \vdots \\ Y_{N1} \ldots Y_{NK} \ldots Y_{NN} \end{matrix} \right\| \left\| \begin{matrix} E_1 \\ \vdots \\ E_K \\ \vdots \\ E_N \end{matrix} \right\| \tag{9}$$

The matrixes of equation 2 may be compounded as shown and the positions of the variables I_K and E_K interchanged with the result that

$$\left\| \begin{matrix} I_1 \\ \cdot \\ \cdot \\ \cdot \\ E_K \\ \cdot \\ \cdot \\ \cdot \\ I_N \end{matrix} \right\| = \left\| \begin{matrix} Y_{11} - Y_{1K} \; Y_{KK}^{-1} \; Y_{K1} \ldots Y_{1K} \; Y_{KK}^{-1} \ldots Y_{1N} - Y_{1K} \; Y_{KK}^{-1} \; Y_{KN} \\ \cdot \qquad\qquad \cdot \qquad\qquad \cdot \\ \cdot \qquad\qquad \cdot \qquad\qquad \cdot \\ \cdot \qquad\qquad \cdot \qquad\qquad \cdot \\ -Y_{KK}^{-1} \; Y_{K1} \quad \ldots \quad Y_{KK}^{-1} \quad \ldots \quad -Y_{KK}^{-1} \; Y_{KN} \\ \cdot \qquad\qquad \cdot \qquad\qquad \cdot \\ \cdot \qquad\qquad \cdot \qquad\qquad \cdot \\ \cdot \qquad\qquad \cdot \qquad\qquad \cdot \\ Y_{N1} - Y_{NK} \; Y_{KK}^{-1} \; Y_{K1} \ldots Y_{NK} \; Y_{KK}^{-1} \ldots Y_{NN} - Y_{NK} \; Y_{KK}^{-1} \; Y_{KN} \end{matrix} \right\| \left\| \begin{matrix} E_1 \\ \cdot \\ \cdot \\ \cdot \\ I_K \\ \cdot \\ \cdot \\ \cdot \\ E_N \end{matrix} \right\| \tag{10}$$

The only difference in the coefficient matrix of equation 3 and that obtained by the application of steps 1 and 2 of the paper is in the algebraic signs of the elements in the Kth column. The method described in the paper has the effect of manipulating equation 2 into the form of equation 3 with the elements in the Kth column of the coefficient matrix multiplied by -1 and with I_K replaced by $-I_K$.

The positions of another pair of variables can now be interchanged in an identical manner and this process continued step by step until the desired equation 1 is determined. It is obvious that the impedance matrix could also be used as a basis for beginning the computations.

Fig. 2 shows a flow diagram of a program that we have been using to accomplish these computations. The following notation applies to this flow diagram.

K = terminal pair where variables are being interchanged
J, M = indexes of terminal pairs
N = number of terminal pairs
A = index of number of times major loop has been traversed or is in process of being traversed
K_A = value of K corresponding to a particular value of A
B_{JM} = number in position JM at start of major loop for a given value of A
B_{JM}' = number in position JM at completion of major loop for a given value of A

The program described by Fig. 2 has several features worthy of mention:

1. All of the elements of the matrix are retained. The discrepancies between corresponding off-diagonal values are useful as a partial indication of the accumulated round-off error, although other tests are applied to the results to give a more qualitative indication. The program can be easily modified to exploit the symmetries of the matrix.

2. Intermediate results are available at the completion of any cycle of the major loop.

3. The order in which terminal-pairs are treated can be controlled by means of the correspondence established between values of A and K_A. This can produce certain computational advantages and is also useful for producing desired intermediate results.

4. When A_M is set equal to N, the result is a complete inversion of the matrix.

The similarities and differences between the program described here and that described in the paper are obvious. The limitations described in the paper also apply here.

REFERENCE

1. DIGITAL COMPUTATION OF POWER FLOW—SOME NEW ASPECTS, H. W. Hale, R. W. Goodrich. *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 78, 1959 (*Paper no. 59-224*).

W. C. Marble (Philadelphia Electric Company, Philadelphia, Pa.): The authors present a thought-provoking paper: their method of matrix inversion permits a reduction in storabe requirements and calculating time which will be valuable for digital computation. One wishes, however, that they had been less laconic in their description of the method. My first attempt to apply the method produced a technique which successfully inverted strongly diagonal matrices, but failed to invert a simple, well-conditioned matrix with a weak diagonal. This resulted from a misinterpretation of the rules given in the paper. To prevent others from being similarly mislead, a fuller description of the authors' method would be helpful.

The method presented permits one to take advantage of symmetry throughout the inversion process; it also illustrates a fa-

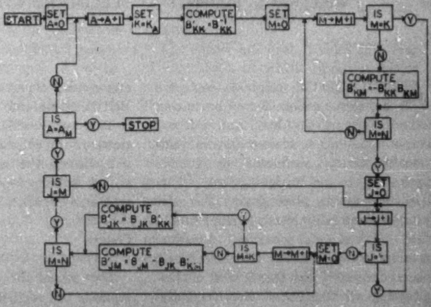


Fig. 2. Flow diagram for interchanging variables

cility in comprehending power system operation in mathematical terms, and vice versa, a characteristic of Mr. Shipley's work which should be admired, and imitated.

R. Bruce Shipley and Dorothy Coleman: The authors wish to express sincere appreciation for the prepared comments. We apologize to Mr. Carter for not having referred to his excellent paper (see reference 1 of the Carter discussion). It is true that the arithmetic is similar if the entire matrix is to be used. However, the use of minus on the pivotal term causes a symmetrical matrix to remain symmetrical throughout all the operations and reduces the memory requirements as well as the arithmetic by approximately one half. This seemingly trivial difference reduces the cost of matrix inversion by one half and doubles the number of matrix elements that can be accommodated in a given digital memory size. A 16,000-word memory will accommodate a symmetrical matrix slightly larger than 100 by 100 in complex numbers, when stored in triangular form. If the complete matrix is to be stored, a 32,000-word machine would be required, and this increases the monthly rental by almost $5,000. Mr. Carter's suggestion of always processing the largest diagonal element first will improve accuracy. The use of the largest element in the matrix as a pivotal element should be expected to improve accuracy, but if it is placed in the diagonal position by interchanging rows, the matrix does not remain symmetrical and the advantage of using only half the matrix is lost.

We agree with Mr. Converti that the direct inversion method is faster than the iterative method when the matrix can be stored in high-speed memory. We are not sure that it will be faster than Mr. Converti's iterative method[1] if the matrix must be stored on tape. Neither are we sure that the direct method is superior to the iterative method if roundoff error becomes a serious problem. To date we have had no serious problem with roundoff errors in inverting power system matrices. We have used the equivalent circuit method suggested by Mr. Converti for about one year and find it very satisfactory from the standpoint of speed and accuracy. It might be noted that equivalent circuits can be derived from the impedance matrix as well as from the admittance matrix. This requires the inversion of only that part of the impedance matrix representing the busses to be held, while the remaining portion of the matrix is simply ignored. The reference bus must be one of the points held. We employ a subroutine that interchanges the reference bus with any other bus in the system, enabling us to hold any combination of busses represented by the impedance matrix. The work done by this subroutine would require about 12 seconds for a 100 by 100 complex matrix on an IBM *704*.

The Hale-Goodrich idea of solving load flow with certain sets of currents as variables and other sets of voltages as variables was possibly first applied by Shipley and Hochdorf.[2] The exact technique of Hale and Goodrich was not used because the additional matrix product RI (where R is the real component of the impedance matrix) was needed. It is not apparent that RI is available from the Hale-Goodrich method without additional computations. It is apparent that the subject inversion method will permit a larger power system load flow to be used since it preserves symmetry. Further, it is more convenient since the sets of variables can be chosen at random in the matrix equation. It will be necessary to change the signs of the variables as they are transferred back and forth across the equality sign.

In reply to Mr. Marble's comments, the following numerical example may be used to demonstrate the individual steps as applied to a complete matrix:

Invert the matrix:

$$\begin{matrix} -23 & 11 & 1 \\ 11 & -3 & -2 \\ 1 & -2 & 1 \end{matrix}$$

Applying rules 1 and 2 to the Y_{33} element gives

$$\begin{matrix} -24 & 13 & -1 \\ 13 & -7 & 2 \\ -1 & 2 & -1 \end{matrix}$$

Next, apply rules 1 and 2 to the $Y_{22}(-7)$ element of the new matrix to give

$$\begin{matrix} 0.142833 & 1.857141 & 2.714282 \\ 1.857141 & 0.142857 & 0.285714 \\ 2.714282 & 0.285714 & -0.428572 \end{matrix}$$

Complete the procedure by applying rules 1 and 2 to the Y_{11} (0.142833) element to give minus the inverse

$$\begin{matrix} -7.001183 & -13.002183 & -19.003185 \\ -13.002183 & -24.004030 & -35.005880 \\ -19.003185 & -35.005880 & -52.008575 \end{matrix}$$

This matrix was taken from the book by Hartree,[3] who gives the inverse as

$$\begin{matrix} 7 & 13 & 19 \\ 13 & 24 & 35 \\ 19 & 35 & 52 \end{matrix}$$

REFERENCES

1. APPLICATION OF ROW-BY-ROW MATRIX INVERSION TO POWER SYSTEM PROBLEMS, V. Converti. *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 78, June 1959, pp. 413–16.

2. EXACT ECONOMIC DISPATCH—DIGITAL COMPUTER SOLUTION, R. Bruce Shipley, Martin Hochdorf. *Ibid.*, vol. 75, Dec. 1956, pp. 1147–53.

3. NUMERICAL ANALYSIS, D. R. Hartree. Oxford University Press, London, England, 1952.

A New Approach to Training Telephone Engineers

WILLIAM C. BURNETT
MEMBER AIEE

LEONARD C. ADAMS
ASSOCIATE MEMBER AIEE

A PICTURE OF telephone engineers living in a college dormitory, eating in a college dining hall, and walking across a campus to classes on communications theory conducted by college professors is an intriguing one. This has been taking place during the past year in South Carolina. It is part of a new method of providing technical training to experienced telephone engineers developed by the School of Engineering of Clemson College and the engineering department of the Southern Bell Telepone and Telegraph Company.

This paper will tell how this revolutionary form of training came into being, will discuss the manner in which the school is being operated, and will present an evaluation of the early results.

Origin

This school had its beginning in two separate but related events. The Southern Bell Telephone and Telegraph Company recognized that it needed a survey of the needs for training in its engineering department. To this end a study was made during the summer of 1956 by a college professor, whose assignment was to explore the needs for training telephone company engineers and the training program as it existed, and to make recommendations for a comprehensive training program. Frequent contact was made with the training people of the engineering department, and trips were

---

Paper 58-1152, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958; the Winter General Meeting, New York, N. Y., February 1–6, 1959; and the Southeast and South Central Meeting, Atlanta, Ga., April 8–10, 1959. Manuscript submitted June 16, 1958; made available for printing June 24, 1958.

WILLIAM C. BURNETT is with the Southern Bell Telephone and Telegraph Company, Birmingham, Ala. LEONARD C. ADAMS is with Clemson College, Clemson, S. C.

made into the operating areas to interview engineers at all levels. The study revealed that the company had a shortage of engineers who might be available as instructors in telephone training courses. As a result, one suggestion was that it might be possible to make use of professional instruction, that is, teachers from college engineering departments.

During the same summer, the transmission engineers of the company met to discuss their needs, and found that one of their greatest was specialized training in telephone transmission engineering for their people. When this was discussed with the personnel of the training section, it was pointed out that prior commitments would make it impossible for a transmission engineering school to be conducted before 1958. The transmission engineers were completely dissatisfied with this situation. The need for early training was urgent.

There were two reasons for this urgency. It had been many years since the company had conducted any sort of transmission school, and during this time many new people had come into the transmission sections. But of greater importance was the expected effect of the vast changes that are taking place in the telephone communications industry. In very recent years the industry has begun to provide a multitude of new types and kinds of services that people and business want and need in this rapidly changing world. In the past the Bell Telephone System has generally followed a system of engineering that begins in the Bell Telephone Laboratories, where basic research is done. General application engineering is a major activity of the operation and engineering department of the American Telephone & Telegraph Company. It is here that Bell System engineering practices are prepared. These are distributed to all the operating companies which, in turn, use them in their local application engineering. When special problems arose, or new and different communications services were proposed, these matters were frequently referred by local area engineers to their headquarters and often to the operations and engineering department in New York City.

Recently the Southern Bell Company decentralized practically all engineering to the nine state areas, which now develop solutions to the special problems and design facilities to care for new and different services. As a consequence the work of the transmission engineers in the state areas has become much more complex and carries far more responsibility. Therefore, their staffs require more thorough training than before.

Another consideration was the nature of the training which the transmission engineers desired for their staffs. From the late 1920's through the 1940's the Southern Bell transmission schools had combined theory with application, making use of equipment and circuits that were in common use throughout the Bell System. Students came from both engineering and operating departments, and the engineering emphasis was on application rather than fundamental principles. Recognizing that basic knowledge is of greater value to the design engineer than is application knowledge, the transmission engineers wanted training to stress the underlying theories of communication circuit design. It was felt that not only would the basic theory be remembered better and longer, but also it would better equip an engineer to tackle and solve the problems of providing new and different communications services.

Further discussion by the transmission and training groups resulted in a proposal to combine the need for early training with a suggestion that college instructors present courses for a telephone school. Since much of the training desired would be basic and of a theoretical nature, it was clear that professional instructors could do this teaching better than telephone engineers.

Overtures were made to Clemson College in South Carolina, and a favorable response was received. Then the transmission engineers established a steering committee to organize the broad aspects of the training course desired, and an outline was presented to the college. Details were developed in joint discussions between the college staff and telephone company engineers, and both groups became enthusiastic about the possibilities of operating a school of the type desired.

The college regards this program of instruction as an industrially supported engineering extension service. Clemson has, for many years, conducted short-courses for agricultural groups. Thus the administration readily accepted the program. This was an opportunity to serve southern industry and, at the same time, to bring about an association with benefits accruing to both company and college. Final acceptance of the idea rested with the electrical engineering faculty. This group adopted the program in a formal meeting without a dissenting vote. Five members of the staff requested a part in the project and others offered to serve if needed.

Advantages of Campus-Type School

A major advantage of the Clemson plan is the use of professional teaching. It is logical that those persons experienced in teaching will do a better job than those whose life work is applied engineering. The ability of a man to design or to solve engineering problems does not necessarily equip him to explain them. Thus, many excellent engineers are poor teachers. Furthermore, it is felt that company instructors could not cover in the allotted time as much ground as do the professors; the company instructors wander off the highway of theory onto the interesting byways of telephone application and personal experiences.

Another major advantage comes from having students in an environment that is conducive to study and learning. Thus, housing telephone students in a domitory, feeding them in a college dining hall, and assembling them in regular classrooms provide atmosphere and climate that are in very favorable contrast to the former arrangement of holding schools in city hotels, which are inherently surrounded by interesting but time-consuming diversions.

Further, there accrues a net saving in engineering manpower because professional teachers can teach fundamentals with less added work than company instructors.

Course of Instruction

In August 1956, the company's transmission engineers chose a steering committee composed of five transmission engineers with the engineering training supervisor as an ex officio member.

From the beginning it was intended that this school would be basically fundamental. As stated previously, it was believed that in the long run engineers would obtain the most benefit from training in fundamentals. Consequently theory, rather than application, has been emphasized.

The school has been divided into four courses, each of which covers a major subject or grouping of subjects. The first course ran 3 weeks; each of the other three are 2 weeks in length. Each course is repeated as many times as needed to accommodate the number of students to be sent. The company regards this arrangement as superior to previous similar schools which the student attended continuously for approximately 8 weeks.

In the first place, this plan promotes efficiency in that once the teaching staff has prepared work for a course, it can present the material every 2, 3, or 4 weeks with a minimum amount of time for review of the lesson plan. Second, the absence of an engineer from his desk for periods of 2 weeks each is far less disrupting to production than is a single absence of 9 weeks. Also, most of the men are much happier if they are not away from home for extended periods. Finally, the divided plan provides desirable flexibility in scheduling the students for those courses which they most need. Some men will take all courses while others may require only one.

The first course, 3 weeks, covers basic transmission theory, including acoustics, transmission lines, and active elements. The second course, 2 weeks, deals with the theory of low-frequency and voice-frequency induction, protection, and corrosion. In the third course, also 2 weeks, the subjects are crosstalk and carrier transmission. The final 2-week course embraces the fundamentals of radio. An outline of the subjects is given in the Appendix.

It was mentioned previously that this school is charged with teaching the basic theory of telephone communications. The application training which should follow has been made a responsibility of the nine state areas. The depth and breadth of this application training will vary according to the needs of the engineers in each area.

Using material furnished by transmission engineers, the professors prepared their own lectures and lesson plans. Quite naturally the nature of material furnished by the company varied with the subject. For instance, the subjects of acoustics, transmission lines, and active networks are generally familiar to an electrical engineering teaching staff. The steering committee and representatives of the college developed the course outline. The individual professors in the electrical engineering department then completed the preparation.

Without recourse to telephone material or reference to telephone company problems in this field, a professor in the chemistry department prepared and taught an excellent course on the theory of corrosion. On the other hand, telephone engineers who are specialists in these subjects prepared detailed lectures for the course on crosstalk theory, transposition theory, and transposition design.

Because of our firm conviction that the quality of teaching by professionals is very much the better, to the maximum practical extent teaching is being done by members of the college faculty. However, where the subject is completely unfamiliar to the faculty, telephone engineers who are considered to be suitable teachers take over the assignment. Most theory is taken by the college faculty, while those applicaton subjects that are peculiar to the communications business become the responsibility of telephone engineers. About 85% of the classroom hours are taught by the professors.

The amount of time spent in classroom work is based upon the schedule of a typical undergraduate engineering student. This amounts to about 24 contact hours per week. Although the volume of subject matter covered is considerable, it is not deemed practical from any point of view to hold the students in formal session for more than 20 to 24 hours each week. Furthermore, homework reading and problem assignments require from 1 to 2 hours out-of-class study per contact hour.

A course consists of several independent subjects each of which is assigned to a different teacher. Thus, the overload on any one staff member is held to a safe minimum. Once the subject breakdown is made, scheduling the program is no great problem. Participants indicate the hours available and their preference of hours for subject offerings. As a result, the schedule literally adjusts itself to an even spread throughout the week.

Most of the textbooks being used are standard technical books. Some are provided on the basis of one for each student, whereas others are made available in several copies for reference use. The texts have been supplemented by Bell System Practices, Bell Telephone Monographs, and other suitable material.

Usually the teacher furnishes to the students handouts consisting of lesson plans and other pertinent material. These promote efficiency in teaching and learning, and provide references for the engineers after they return to their jobs.

In the References is a list of the textbooks which are being used in the school, not including The Bell System material. The steering committee has recommended that the transmission sections in the state areas build up their technical libraries with many of these texts.

In the interest of conserving time, no laboratory work is included in the course but demonstrations are used to emphasize certain theory.

The faculty instructors feel that it is not fair either to the short-course student or to the college to grade the men taking the course. They are not on the campus long enough for this. However, the company wishes to know which of its engineers excel in fundamental theory, so two grades are given. All participants receive a grade of S (satisfactory) except a small percentage who do exceptionally good work. These receive a grade of E (excellent). The company is furnished no further grading information. Grades are based upon homework papers and a series of quizzes given during the course.

Students

The subject material has been pitched at the college senior and graduate level. Therefore the general student qualification is that he should be an electrical engineering graduate or possess equivalent theoretical background. Equivalence could be attained by a college or high school graduate who had considerable technical experience, informal education, and training within the company.

The original intent was to provide advanced training for transmission engineers and for engineers who were good prospects for transmission work. Then it was decided to open the door to transmission maintenance personnel from the plant department.

No tests are given to prospective students. Rather, they are chosen by the state areas on the basis of their formal education, company training, telephone experience, present duties, and ability to absorb high-level training. We believe that generally good choices have been made.

Several weeks prior to his attendance at a semester of schooling, each student is sent a textbook and other material for advance study. This enables him to determine what reviewing he needs to do and to acquire some familiarity with the text. In turn, this promotes efficiency in learning while he is at Clemson.

Mechanics

Telephone Company students are housed in a dormitory that has always been used for short-course groups. During the last 5 years, Clemson has erected new housing which has been more than adequate for her growing body of about 4,000 resident students. Thus the housing officer has made a block of 25 rooms available to the telephone school for its duration, or until such time as they are needed for the student body.

Clemson students eat, family style, in a single tremendous dining hall. Normally, a boarding student pays for room and board at the beginning of the semester and is admitted to the dining room upon identification by card. Short-course students are accommodated in the same way. The cost per student is slightly greater to allow for maid and linen service, a service not available to the undergraduate resident.

Evaluation

Student reaction must be considered in the light of the intent of the school, which is to provide high-level training in the basic theories of communications circuits and components. Many of the students have come away from a 2-week or a 3-week session feeling that they had been swept up onto a cloud and left there, unable to pierce the fogs and mists of theory and see the everyday job back on earth. This is neither an unexpected development nor one that causes us concern. The planners of the Clemson school wanted major emphasis on basic theory and intended that the state areas would carry out suitable application training. As an example of this, the Alabama area will hold, upon the completion of the current semester of the Clemson school, a 3-day seminar in which applications will be discussed. This seminar will be attended by all of the Alabama engineers who have been enrolled in the Clemson program. By this means the training of engineers will be made more nearly complete.

After a good look at the Southern Bell curriculum and schedule at its Clemson school, and with the usual engineering curriculum in his mind, one professor observed, "It looks like Southern Bell's intent is to whet the appetite of its students." He was contrasting the very rapid coverage in the telephone course with the thorough subject exploration and exposition that is typical of college courses. Having in mind the practical aspects of time limitations and the expected high quality of students, the steering committee planned a course of study that covered a great deal of ground quickly. We hope that we are not wrong in expecting that this communications training will stir up in each man such great interest that he will turn to a deeper home study of many of the subjects he encountered at Clemson. Certainly, an engineer who wishes to grow and develop and succeed will feel strongly that he needs to exert off-the-job effort to further his own development. There is, of course, opportunity to sandwich into the day-by-day job reading and study of technical material, but the man who is sincerely ambitious will realize that 8:30 to 5 efforts are not enough.

We expect that engineers who have had this training will have been equipped to get for themselves the most out of the technical information normally available for their use. And when new problems arise, problems that have not been anticipated or problems concerning new and different communications' services, the transmission engineers will have the basic knowledge to solve these problems. This is vastly important in the day of decentralization of telephone engineering and of a host of new and different services wanted and offered.

Highly important is the fact that this program is bringing about a new and closer relationship between education and industry. Formerly, industry has been considered to be interested only in profit, and the college to be interested only in courses in theory. The Clemson venture shows that this need not be so, for participation in this joint activity has developed a mutual interest, and a realization that both education and industry can benefit from working together.

Conclusion

When in 1956 the Southern Bell Telephone Company recognized the great need for technical training for its transmission engineers, it sought ways of providing this training. Full study of the problem developed a firm conclusion that basic training in theory and principles would be of greater value and more enduring than study of systems and applications.

When prior commitment of the engineering department's training staff indicated a serious delay in starting the needed training, inquiry located a state college that was interested in conducting courses of study on its own campus.

The development of a working arrangement with Clemson College was a most happy solution in that classes could begin soon, the teaching of basic theories could be admirably done by college professors, professional teachers can teach better than engineers, and the holding of classes on a campus would provide a climate highly conducive to learning.

By these means we are providing training that should enable our transmission engineers to think through unusual circumstances not covered in our Bell System Practices.

Appendix. Outline of School

Course I. (3 weeks)

A. Acoustics
   1. Sound
   2. Acoustical engineering and noise
   3. Information theory

B. Transmission lines
   1. Network theory, passive networks
   2. Electric wave filters
   3. Transmission line theory

C. Active elements
   1. Vacuum tubes
   2. Transistors
   3. Gain devices

Course II. (2 weeks)

A. Noise-frequency induction
   1. Fundamentals, noise sources, components
   2. Influence, harmonics, measurements
   3. Coupling and shielding
   4. Susceptiveness, investigations, surveys

B. Selective devices
   1. Theory, shunts, capacitors
   2. Propagation of harmonics

C. Low-frequency induction
   1. Sources
   2. Symmetrical components
   3. Power system faults
   4. Coupling factors
   5. Shielding

D. Protection
   1. Lightning
   2. Telephone plant
   3. Protective measures

E. Corrosion
   1. Theory

Course III. (2 weeks)

A. Crosstalk
   1. Theory
   2. Types of crosstalk
   3. Toll cable crosstalk
   4. Number splicing
   5. Exchange cable crosstalk
   6. Open wire crosstalk
   7. Carrier system co-ordination

B. Carrier transmission
   1. History, why modulate
   2. Kinds of modulation
   3. Sampling principle
   4. Quantization, code transmission, multiplexing
   5. Amplitude modulation
   6. Frequency modulation
   7. Pulse modulation
   8. Oscillators
   9. Linear amplifiers
   10. Filters
   11. Compandors
   12. Noise
   13. Systems
   14. Co-ordination of systems
   15. A look into the future

Course IV. (2 weeks)

A. Radio
   1. Modulation
   2. Oscillators
   3. Amplifiers
   4. Discriminators
   5. Limiters
   6. Waveguides, coaxial transmission lines
   7. Antennae
   8. Wave propagation
   9. Noise
   10. Air/ground
   11. Tropospheric or scatter

B. Theory of probability

References

Textbooks

1. Active Networks, V. C. Rideout. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1954.

2. Applied Electronics, T. S. Gray. John Wiley & Sons, Inc., New York, N. Y., second edition, 1954.

3. Communication Engineering, W. L. Everitt, G. E. Anner. McGraw-Hill Book Company, Inc., New York, N. Y., third edition, 1956.

4. Electric Transmission Lines, H. H. Skilling. McGraw-Hill Book Company, Inc., 1951.

5. Electrical Communication, A. L. Albert. John Wiley & Sons, Inc., third edition, 1950.

6. Electrical Transmission and Distribution Reference Book. Westinghouse Electric Corporation, East Pittsburgh, Pa., 1950.

7. Electronic Engineering, Samuel Seely. McGraw-Hill Book Company, Inc., 1956.

8. Electron-tube Circuits, Samuel Seely. McGraw-Hill Book Company, Inc., second edition, 1958.

9. Acoustical Engineering, H. F. Olson. D. Van Nostrand Company, Inc., Princeton, N. J., third edition, 1957.

10. Elements of Pulse Circuits, F. J. M. Farley John Wiley & Sons, Inc., 1955.

11. Frequency Analysis Modulation and Noise, Stanford Goldman. McGraw-Hill Book Company, Inc., first edition, 1948.

12. Modulation Theory, H. S. Black. D. Van Nostrand Company, Inc., 1953.

13. Pulse and Digital Circuits, Jacob Millman, H. Taub. McGraw-Hill Book Company, Inc., 1957.

14. Electronic and Radio Engineering, F. E. Terman, et al. McGraw-Hill Book Company, Inc., fourth edition, 1955.

15. Symposium on Corrosion Fundamentals, edited by A. de S. Brasunas, E. E. Stansbury. University of Tennessee Press, Knoxville, Tenn., 1956.

16. Theory and Application of Microwaves, A. B. Bronwell, R. E. Beam. McGraw-Hill Book Company, Inc., 1947.

17. Transistor Electronics, A. W. Lo, et al. Prentice-Hall, Inc., 1955.

18. University Physics, F. W. Sears, M. W. Zemansky. Addison-Wesley Publishing Company, Reading, Mass., second edition, 1956.

19. Vacuum Tube Circuits and Transistors, L. B. Arguimbau, with R. B. Adler. John Wiley & Sons, Inc., 1956.

Technical Article

1. Theoretical Limitations on the Rate of Transmission of Information, W. G. Tuller. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 37, no. 5, May 1949, p. 468.

Discussion

Ellis F. King (University of California, Los Angeles, Calif): The authors' reference to emphasis on "underlying theories of communication circuit design" is noted with interest. This is in contrast to an "appreciation" course (to employ the usual telephone terminology), an experience roughly correspounding to a "survey" course as the term is used in academic circles.

Appreciation courses are frequently characterized by their concentration within a few days, a relatively large number of hours per day, no outside assignments or study time, and the lack of any real increase in the student's ability to apply fundamental principles to the solution of engineering problems. Properly presented, they do provide an improved conversational knowledge of the "state of the art" and tend to promote better understanding of general company operations on an interdivisional and interdepartmental basis. Clearly, these results have merit. No apology need be made for such courses if company objectives are consistent with the limitations of such an educational experience.

In contrast, so-called theoretical courses emphasize principles rather than techniques, concepts rather than precepts, and ideas rather than hardware. Competently administered, they provide the framework of technical competence for performance at a professional engineeering level. There is rather general agreement that the anatomy of such courses should include substantial outside study time, problem assignments, and examination procedures.

It is most important in any industrial schooling that management distinguish between these fundamentally different sets of objectives and provide the proper educational format for the desired results.

Mr. Burnett and Mr. Adams have reported teaching procedures which, within the time limitations extant, appear admirably consistent with their stated objectives.

Further comment on two aspects of the paper seem in order.

1. The *a priori* assumption of satisfactory or excellent performance.
2. The designation of the program as a "training" experience.

It is conceivable that personnel might participate in such a program with performance which would cause supervision to question the advisability of continuation and/or participation in possible future courses based on the earlier work. We suggest a third category; persons for whom further consideration may be desirable in the form of consultations between the individual and his supervision. It is felt that such a procedure offers the possibility of improved communication in relating employee development to long-range company and employee interests, as well as better information for future selections. Details may be handled in a manner calculated to avoid embarrassment on the part of anyone concerned.

Finally, we would greatly prefer that such programs as the authors describe be designated as education rather than training. To many, and particularly to the academically oriented, training is associated with enhancement of ability at the craft and technician level basic theory, and how to perform in a particular job classification, are two different matters. Less confusion, both with higher management internally and the profession externally, would accompany this suggested exercise in semantics.

W. J. Miller (University of Alabama, University, Ala.): The sponsors of this educational program are to be congratulated on their ability to recognize that tailors are superior to hatters in making a coat. In other words, they have seen that proper teaching of principles and theories requires experienced teachers, and cannot be done effectively by practicing engineers.

They have also demonstrated an understanding that an engineering teacher rarely knows applications and practice sufficiently well to present such matters effectively.

This understanding and use of basic facts makes this program more promising than the usual short course presented to upgrade personnel.

There is usually some confusion in the use of the terms education and training. Since this course was not given to produce proficiency in the use of some new device or method, it appears to be a program in education as contrasted with training.

Since all individuals taking this course have backgrounds of Bachelor's level in electrical engineering, the material presented in this course is partly "refresher" and partly post-graduate in nature. Thus, this program appears to be, to a certain extent, following the pattern of the so-called post-graduate courses of the medical profession. This may indicate that the engineering profession is coming of age and an educational plan of this type is to be highly commended. It may become a pattern for professional improvement in other industries.

Vol.
Pt.
195

Analysis and Design of a Transistor Linear-Delay Circuit

R. P. NANAVATI
NONMEMBER AIEE

Synopsis: In this paper an analysis is presented of the emitter-coupled linear-time-delay circuit as well as a systematic design procedure which starts with a practical set of specifications. Some important fundamental limitations on the operation and design of the circuit are also presented. Experimental data are included to give an idea of the range of validity of the theory. A range of the ratio of maximum-to-minimum pulse width as high as several hundred has been obtained as compared with 10 or 20 obtainable from a similar vacuum-tube circuit. A ratio in excess of 30 was obtained with a deviation from linearity of less than 3%. Stages of this circuit have been successfully operated in cascade.

FIGURE 1 shows the delay circuit chosen from among many single-shot multivibrators. It might be noticed that the normal coupling from the collector of T_2 to the base of T_1 found in other monostable circuits is missing. For this reason the base of T_1 is an ideal place to trigger the circuit, and the collector of T_2 is an ideal place from which to obtain the output signal. The most important reason for the choice of this circuit is that its output pulse duration turns out to be an accurately linear function of the base voltage of T_1.

Considering the circuit performance as piecewise linear and the transistor to be represented as a dependent current generator, and neglecting the voltage drop between the emitter and base of a conducting transistor, it can be shown quite readily that the output pulse width is given by

$$t_0 = T\ln\left(1-\frac{V_{e0}}{E_{cc}}\right) + T\ln\left(1+\frac{\beta_1 R_1 V_{b11}}{(R_8+\beta_1 R_7)(E_{cc}-V_{e0})}\right) \quad (1)$$

where the notation for V's is as follows: The first subscript denotes at which terminal (i.e., emitter, base, or collector) the voltage is measured; the second subscript denotes that the voltage in question is measured on T_1 if the subscript is 1 and T_2 if the subscript is 2. The third subscript denotes the voltage before trigger is applied if the subscript is 0, and after the trigger if the subscript is 1. Because the emitters of T_1 and T_2 are at the same potential, their voltage is simply denoted by V_{e0}, V_{e1}.

In deriving equation 1 it has been assumed that

$$\frac{\beta_1 R_7 V_{b11}}{(R_1+\beta_1 R_7)E_{cc}} \ll 1$$

which is found to be true for most practical values obtained. Equation 1 can be approximated by

$$t_0 = T\ln\left(1-\frac{V_{e0}}{E_{cc}}\right) + T\frac{\beta_1 R_1 V_{b11}}{(R_8+\beta_1 R_7)(E_{cc}-V_{e0})} \quad (2)$$

and

$$T = R_6 C \quad (3)$$

if the higher order terms in the expansion of the logarithmic term are negligible as compared with the ones in equation 2. For most practical parameter values this is normally true. Under these conditions the output pulse width is seen to be a linear function of the input voltage. In addition, it should be noticed that if

$$R_8 >> \frac{R_4 R_5}{R_5+R_4} \quad (4)$$

then V_{b11} can be conveniently replaced by V_{b10} in equation 2 so that the output pulse width can be expressed wholly in terms of quiescent quantities. In the interest of linearity the foregoing inequality is made to satisfy.

Establishing a Limit on the Minimum Pulse Width Obtainable

Even though one of the advantages of the circuit of Fig. 1 is the linearity of performance, one of the important figures of merit is the ratio of the maximum-to-minimum output pulse widths. The circuit performance is not linear over the entire range from maximum-to-minimum obtainable pulse widths. Just the same, by appropriately calibrating the input potentiometer the circuit can be operated over a considerable range of output pulse widths.

Fig. 2 shows the base and collector waveforms of T_1 and T_2. In Fig. 2(C) V_{b2c0} stands for the base voltage that just cuts off base current to T_2. A careful study of the waveforms indicates that the following is true.

$$t_{0\,\mathrm{min}} \geq t_{d2}+t_{r2}+t_{s2} \quad (5)$$

Another fundamental limit which is obtained from the waveforms is the delay in the formation of output pulse width which is given by

$$t_{\mathrm{delay}} = t_{r1\,\mathrm{off}}+t_{s1}+t_{s2} \leq t_{r1}+t_{s1}+t_{s2} \quad (6)$$

Equations 5 and 6 are absolute limits. The rise, t_r, saturation, t_s, and decay, t_d, times for the transistor themselves have been worked out by Ebers and Moll[1,2] and more recently by Easley[3] taking into account the collector capacitance and load resistance. The decay time has been discussed in detail in a recent paper.[4]

Maximum Trigger Rate Considerations

It is possible to solve (through a straightforward though tedious algebra) for the frequency at which the real part of the impedance that the capacitance c in Fig. 1 sees ceases to be negative. This frequency has a range depending on the value of the external circuit components and is given by equation 7.

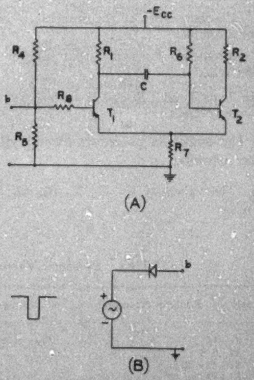


Fig. 1. Common-emitter linear-delay circuit with the simplest coupling network

---

Paper 59-563, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Empire District Meeting, Syracuse, N. Y., April 29–May 1, 1959. Manuscript submitted March 6, 1958; made available for printing February 19, 1959.

R. P. NANAVATI is with Syracuse University, Syracuse, N. Y.

This paper presents the results of one phase of research sponsored by Rome Air Development Center under Contract no. *AF 30(602)-1626.*

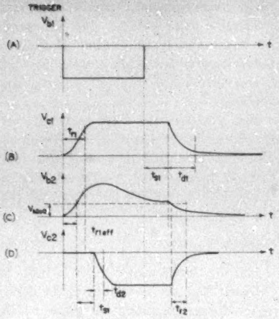


Fig. 2. Waveforms at different points in the circuit of Fig. 3

$$\omega_\beta \sqrt{1+\frac{\beta_0}{\alpha_0}} > \omega_{\max} > \omega_\beta \sqrt{1+\frac{\beta_0 R_7}{\alpha_0 R_1}} \qquad (7)$$

Equation 5 also puts a limit on the maximum trigger rate. Thus the maximum trigger rate is given by

$$f_{\max} = \min\left[f_\beta \sqrt{1+\frac{\beta_0}{\alpha_0}}, \frac{1}{t_{d2}+t_{r2}+t_{02}} \right] \qquad (8)$$

Design Procedure

The problem is to design the monostable circuit subject to the following specifications:

1. A range of available E_{cc}.
2. Maximum expected noise (spurious signals) E_n.
3. Minimum signal voltage available E_s.
4. Minimum output voltage required V_{omn}.
5. T_2 and, if possible, T_1 should not saturate.
6. Characteristics of the transistors used are available.
7. Maximum desirable output pulse duration $t_{0\,\max}$.

(In items 2 and 3 E_n and E_s are peak values of the voltages.)

The synthesis procedure should yield the following.

1. The values of R_1, R_2, R_4, R_5, R_6, R_7, R_8, C.
2. The required characteristics of the diode used.
3. The range of linearity.
4. The maximum ratio of maximum-to-minimum pulse widths.

The following are the steps taken to determine the foregoing quantities:

1. Choose the highest available E_{cc} consistent with the transistor ratings. This gives the best linearity, all other things being equal.

2. Choose a reasonable value of I_2 keeping in mind the fact that $I_2 > I_1$ (where I_2 and I_1 are the emitter currents of T_2 and T_1 respectively, and the α and β for both transistors are assumed identical). This means that the value of I_2 should not be very low, otherwise an unreasonable value of I_1 will have to be chosen.

3. Then choose

$$R_2 \geq V_{omn}$$

(This meets the output requirements.)

4. Choose

$$I_2 R_6 > \beta R_2$$

(This ensures the prevention of saturation of T_2.)

5. Choose

$$R_7 = \alpha \frac{E_{cc}}{I_2} - \frac{R_6}{\beta}$$

This is the analytical derivation of R_7 assuming the transistor characteristics to be piecewise linear.

A more realistic calculation of R_7 may be done graphically in the following way using the transistor characteristics furnished by the manufacturer together with the following equation:

$$R_7 = \frac{\alpha}{I_2}(E_{cc} - I_{b2} R_6)$$

Having located a desired operating point one reads off I_{b2} and I_2 directly from the transistor characteristics. Then having calculated R_7 from the foregoing equation one draws the load line of slope

$$-\frac{1}{R_7+R_2}$$

If the load line does not pass through the desired operating point, one chooses another point and goes through the cycle again until a decent operating point is arrived at. No more than three trips should be necessary.

Table I. Parameter Values Used in the Experimental Circuit

Resistors, Kilohms	Transistors*	Voltage	Capacitors	Diode
$R_1 = 18$	T_1	$E_{CC} = 10$ volts	$c = 0.01$ microfarad	$R_F = 200$ ohms
$R_2 = 6.8$	T_2			$R_B = 500$ kilohms
$R_4 + R_5 = 200$				
$R_6 = 270$				
$R_7 = 2.2$				
$R_8 = 270$				

* General Electric type *2N123*.

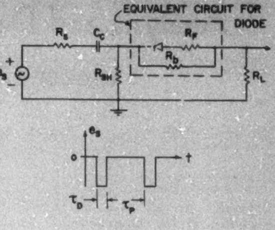


Fig. 3. A typical coupling circuit feeding a load R_L

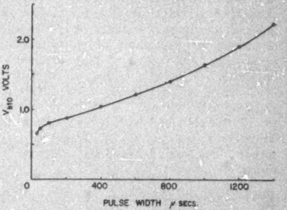


Fig. 4. Linearity data

$E_{cc} = -10$ volts

6. The choice of R_7 sets limits on V_{b10} as follows:

$$|E_s| \geq -|V_{b10}| + |I_2 R_7| \geq |E_n|$$

$$|V_{b10\,\min}| \geq |I_2 R_7| - |E_s|$$

$$|V_{b10\,\max}| \leq |I_2 R_7| - |E_n|$$

Observe that $|E_s| > |E_n|$.

7. Thus the range of voltages which the potentiometer must accomodate is

$$|V_{b10\,\min}| = |I_2 R_7| - |E_s|$$

$$|V_{b10\,\max}| = |I_2 R_7| - |E_n|$$

8. Choose

$$|I_1| < \left| \frac{V_{b10\,\min}}{R_7} \right|$$

where it has been assumed that

$$V_{b10\,\min} = V_{b11\,\min}$$

(This condition ensures the conduction of T_1 during the pulsing cycle.)

9. Choose

Table II. Comparison of Measured and Theoretically Calculated Values

Measured, Microseconds	Calculated, Microseconds	% Error
50	52.7	−5.12
100	105	−4.75
250	240	+4.17
300	281	+6.78

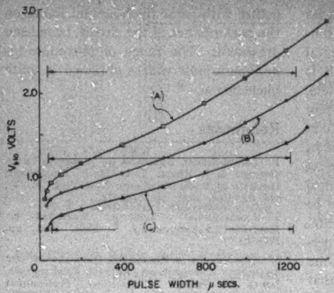


Fig. 5. Linearity data of the output stage of two cascaded stages

A—Ratio = 31, deviation <3%, jitter <0.5%, $E_{cc} = -13$ volts
B—Deviation <3.6%, ratio = 30.5, $E_{cc} = -10$ volts
C—Deviation <4.5%, ratio = 22, $E_{cc} = -7$ volts

$$R_1 < \left| \frac{E_{cc}}{I_1} - R_7 \right|$$

so as to prevent saturation of T_1 during pulsing time. It should be noticed that this will not prevent saturation caused by the trigger source.

10. Choose R_8 so that it satisfies the following inequalities:

$$\frac{R_1 R_7(\beta-1)}{R_7+(1-\alpha)R_1} > R_8 >> \beta R_7$$

The former part of the inequality satisfies the negative resistance criterion (that the circuit must satisfy to be a multivibrator at all) and the latter part of the inequality simulates a current drive at the base of T_1.

11. If R_b is the back-resistance of the diode used to decouple the trigger source from the circuit, then

$$\frac{(R_8+\beta R_7)R_4 R_5}{R_4 R_5+(R_4+R_5)(R_8+\beta R_7)} << R_b$$

or if

$$(R_8+\beta R_7) >> \frac{R_4 R_5}{R_4+R_5}$$

(which must be true if the assumption $V_{b10} \doteq V_{b11}$ is justifiable), then

$$\frac{R_4 R_5}{R_4+R_5} << R_b$$

Table III. Experimental Data for Fig. 6

Minimum Pulse Width, Microseconds	Maximum Pulse Width, Microseconds	% Jitter	C, Microfarads	Maximum Ratio
5	1,800	$\frac{5.6}{200}=2.8$	0.0147	360
5	1,400	$\frac{3.2}{170}=1.88$	0.01	280
5	1,070	$\frac{1.2}{100}=1.2$	0.0047	214
5	365	$\frac{0.7}{100}=0.7$	0.0022	73
5	120	$\frac{0.4}{100}=0.4$	0.0010	24
5	60	$\frac{0.1}{45}=0.22$	0.0005	12

This is one condition that the choice of the potentiometer (R_4+R_5) must satisfy. There is still one more degree of freedom left in the choise of (R_4+R_5). This other degree of freedom may be decided from among a group of practical considerations such as drain on the power supplies, available ohmic values of the potentiometer, wattage dissipation that can be withstood, etc.

12. When

$$I_2 R_7 = V_{b10} \tag{9}$$

T_1 starts to conduct in the absence of trigger. This is the point at which the output pulse width should start to diminish. Hence the maximum pulse width is obtained by evaluating equation 1 with the use of equation 9. Then using specification 7, for the value of the capacitance C

$$C = \frac{t_{0\,\max}}{R_6 \ln\left(1-\frac{V_{e0}}{E_{cc}}\right) + \ln\left[1+\frac{\beta R_1 I_2 R_7}{(R_8+\beta R_7)(E_{cc}-V_{e0})}\right]}$$

This completes the design procedure for the circuit itself.

Cascade Operation Considerations

Fig. 1(B) shows the simplest and the most desirable coupling circuit to be used for single-stage operation, as it offers the fewest restrictions on the circuit performance. If the source cannot tolerate direct current through it, then the circuit of Fig. 3 should be used.

If τ_c and τ_n denote the time constants of this circuit when the diode is conducting and when the diode is not conducting, respectively, the circuit parameters should be so chosen that $\tau_c > 10\tau$ (where τ is the time constant of input pulse rise time) *and* $\tau_n << \tau_p$. The former condition prevents too much loss in effective trigger height.[5] The latter condition ensures the return of the circuit to normality before the next trigger pulse arrives.

The case of several stages of cascade can be handled with the same circuit coupling as that of Fig. 3, except that here the output resistance of the preceding stage and the input resistance of the following stage will replace R_g and R_L, respectively. In practice, the presence of the diode was found necessary to make the output pulse width independent of trigger duration and height.

When very long delays are needed they may be obtained by operating the common emitter circuits in cascade. Thus with such an operation the range of delays can be made quite large without introducing appreciable error.

Experimental Correlation

Fig. 4 shows a large linear range of output pulse width with the voltage V_{b10} as predicted by the analysis. Table I shows the value of the components used and Table II shows the correlation between the values of the output pulse width calculated with the use of equation 1 and the values obtained experimentally from a curve like that of Fig. 4. The correlation is as good as can be expected. The explanation for departure from linearity at large and small values of V_{b10} follows readily from theory.

In vacuum-tube circuits of the cathode-coupled linear-time-delay type, ratios of the maximum-to-minimum output pulse

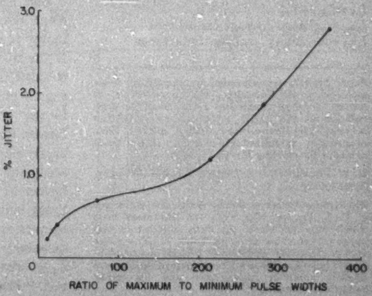


Fig. 6. Variation of jitter with output-pulse-width ratio

width obtainable of 10, or at most 20, have been possible. In the transistor circuit under consideration ratios as high as 800 have been obtained from a single circuit. Restricting deviations from linearity to be within 3%, a ratio as high as 31 has been realized. (Deviation here is defined as the ratio, at any particular point, of the difference between the ordinate of the experimental curve and that of the best fitting linear line divided by the experimental ordinate at the point.) Fig. 4 shows one of these curves.

Fig. 5 shows the output pulse duration as a function of the base voltage V_{b10} for the second stage of a 2-stage cascade circuit where the first stage is triggering the second stage without the use of an auxiliary amplifier.

As Table III indicates, the minimum pulse width obtained from the circuit is independent of the value of C used, whereas the maximum pulse width varies linearly with C. The former fact confirms the analysis that the fundamental limitation on minimum pulse width is imposed largely by device parameters. Thus it would seem possible to obtain large ratios of maximum-to-minimum pulse width limited only by the value of the capacitance used. The maximum value of C is determined in any practical case by the amount of jitter that can be tolerated. Fig. 6 shows an experimental curve of the maximum ratio versus percentage jitter based on the figures from Table III. It might be mentioned that the circuit configuration chosen minimized jitter.

Conclusions

Analysis and detailed design procedure for a transistor linear-delay circuit, together with some fundamental limits on the performance of the circuit, have been presented. The range of maximum-to-minimum pulse widths obtained is quite phenomenal.

References

1. LARGE-SIGNAL BEHAVIOR OF JUNCTION TRANSISTORS, J. J. Ebers, J. L. Moll. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 42, Dec. 1954, pp. 1761–72.
2. LARGE-SIGNAL TRANSIENT RESPONSE OF JUNCTION TRANSISTORS, J. L. Moll. *Ibid.*, pp. 1773–83.
3. THE EFFECT OF COLLECTOR CAPACITY ON THE TRANSIENT RESPONSE OF JUNCTION TRANSISTORS, James W. Easley. *Transactions*, Professional Group on Electron Devices, Institute of Radio Engineers, Jan. 1957, pp. 6–14.
4. TURN ON DELAY TIME AND ITS PREDICTION, R. P. Nanavati, R. A. Johnson. *Proceedings*, National Electronics Conference, Chicago Ill., vol. XIV, 1958, pp. 25–31.
5. PULSE AND DIGITAL CIRCUITS (book), J. Millman, H. Taub. ELECTRICAL AND ELECTRONIC ENGINEERING SERIES, McGraw-Hill Book Company, Inc., New York, N. Y., 1956, p. 34.

An Equivalent Circuit for Transformers in Which Nonlinear Effects Are Present

H. W. LORD
FELLOW AIEE

TEXTBOOKS dealing with the subjects of transformer theory and design generally use the T-equivalent circuit to represent the inductive parameters of a transformer. This paper will point out a limitation of the T-equivalent circuit and will show that the π-equivalent circuit has no such limitation.

Nomenclature

n = transformer turns ratio (secondary turns/primary turns)
L_P = Total primary inductance = $L_{P,l}+L_m/n^2$
$L_S=L_{S,l}+L_m$
$L_{P,l}$ = primary leakage inductance
$L_{S,l}$ = secondary leakage inductance
$L_l=n^2L_{P,l}+L_{S,l}$ = total leakage inductance referred to the secondary
$L_{P,m}$ = magnetizing inductance of primary core portion
$L_{S,m}$ = magnetizing inductance of secondary core portion
L_m = magnetizing inductance of entire core referred to secondary
$\mathcal{R}_P$ = reluctance of primary core portion
$\mathcal{R}_S$ = reluctance of secondary core portion
$\mathcal{R}_l$ = reluctance of leakage flux path

The Well-Known T-Equivalent Circuit

Fig. 1 shows the well-known T-equivalent circuit which has been used for many years to represent a transformer. If the transformer it represents is a close-coupled transformer, such as a conventional public-utility distribution transformer or a small power-supply transformer for electronic circuits, this equivalent circuit is generally adequate for analyzing the performance of the transformer in a system. The total magnetizing inductance, L_m, is determined by an open-circuit test. The primary and secondary leakage inductances, $L_{P,l}$ and $L_{S,l}$, are determined by a short-circuit test. The short-circuit test determines the total leakage inductance, L_l. However, when reflected to the same transformer terminals, one half of this total is arbitrarily assigned to the primary leakage inductance and the other half is assigned to the secondary leakage inductance; i.e., it is assumed that $n^2L_{P,l}=L_{S,l}=L_l/2$. (Since this paper deals mainly with the inductive elements of the equivalent circuit, details for determining the equivalent resistive or loss elements will not be discussed.)

When the value of L_l is only a fraction of L_m, as it is in close-coupled transformers, the error introduced by assuming that $n^2L_{P,l}=L_{S,l}$, even though such equality may not exist, is small. In fact, for such transformers, the equivalent circircuit (neglecting losses) is often simplified to that of Fig. 2 in order to simplify the circuit analysis.

When a transformer involved in an analysis is of the type which designers of power-frequency transformers call high-reactance (high leakage inductance) trans-

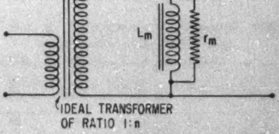


Fig. 1. T-equivalent circuit of a transformer

Paper 59-672, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Baltimore, Md., May 19–21, 1959. Manuscript submitted August 25, 1958; made available for printing March 23, 1959.

H. W. LORD is with the General Electric Company, Schenectady, N. Y.

The author wishes to acknowledge the assistance of P. A. Fessler, who made the laboratory tests which provided all of the data included in this paper except those given in Table I. The latter were provided by a portion of a series of tests conducted in co-operation with engineers of the General Electric specialty transformer and fluorescent ballast departments.

Vol.
Pt.
195

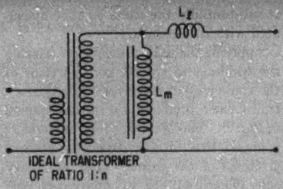


Fig. 2. Simplified L-equivalent circuit

formers and which radio engineers call loose-coupled transformers, the equivalent circuit of Fig. 2 can seldom be used. Moreover, in this type of transformer, if the reluctances of the magnetic circuits associated with the primary and secondary coils differ substantially, the T-equivalent circuit with $n^2L_{P,l}$ and $L_{S,l}$ equal is no longer valid, and an assumption that they are equal leads to substantial errors in an analysis.

When nonlinearities due to core-saturation effects are present, the T-equivalent circuit provides no means for separating the saturation effects associated with the primary-core portions from those associated with the secondary-core portions.

For a low-reactance transformer of normal design, saturation effects have little or no effect on the secondary voltage under various load conditions. However, in these types, the amount of load current and its phase angle with respect to the induced voltage cause the flux densities in certain portions of the core to be different from those in other core portions. For this reason, the T-equivalent circuit is not suitable for analyzing the effects of loading upon flux-density-dependent characteristics of a low-reactance transformer. These effects include harmonic content of the exciting current, core losses, and noise due to magnetostriction.

Saturation effects in the secondary core portions of high-reactance transformers affect both the secondary voltage waveforms and the effective primary-to-secondary voltage ratios. The T-equivalent circuit is practically useless for representing a high-reactance transformer when secondary-core-portion saturation effects are present.

The T-equivalent circuit bears no direct relationship to physical portions of the transformer structure. To many engineers, this may not constitute a serious handicap. But there are probably many others who are disturbed by the lack of direct relationships between the physical properties and dimensions of a device and a circuit which is considered to be its equivalent.

The π-Equivalent Circuit

Fig. 3 shows a π-equivalent circuit for a transformer which is the counterpart of the T-circuit of Fig. 1. The earliest, generally available, published material showing this transformer equivalent circuit is reference 1 which points out that: "Figure 13 is probably more adaptable to the representation of saturation characteristics of the core than Figure 15a." (Fig. 13 of the reference shows a π-equivalent circuit and Fig. 15a shows a T-equivalent circuit.) Actually, as will be clear from the following sections, this quotation is an understatement of the usefulness of this π-equivalent circuit.

If, for the sake of simplicity, the effects of losses are neglected, the π-equivalent circuit reduces to that of Fig. 4.

π-Equivalent Circuit Related to Physical Transformer Structures

One advantage of the π-equivalent circuit is that the magnetic reluctances of certain portions of the transformer-core structure, plus the windings associated with them, determine the values of the inductive elements of the equivalent circuit, and it is not necessary to assume any fixed ratios of their values. Fig. 5 represents the core and coils of a high-reactance transformer. The magnetic reluctances of core portions associated with the primary, secondary, and leakage inductances are labeled $\mathcal{R}_P$, $\mathcal{R}_S$, and $\mathcal{R}_l$ respectively. The reluctance $\mathcal{R}_P$ of the indicated core portion and the turns of the primary coil are factors which determine the value of primary magnetizing

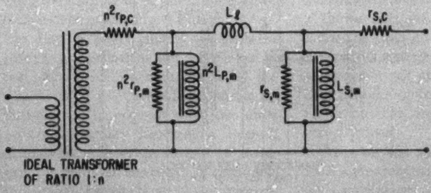


Fig. 3. π-equivalent circuit of a transformer

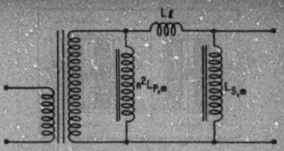


Fig. 4. π-equivalent circuit simplified by neglecting losses

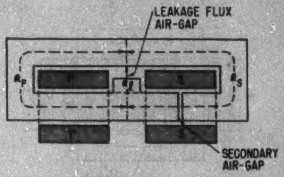


Fig. 5. Core and coil arrangement of a core-type high-reactance transformer

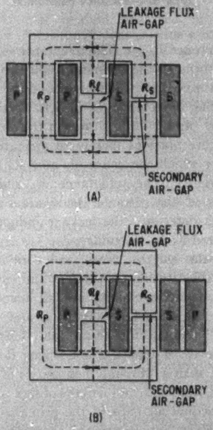


Fig. 6. Two coil configurations on the same core configuration of a core-type high-reactance transformer

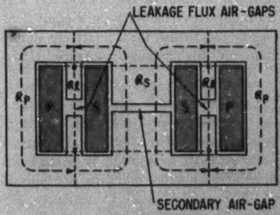


Fig. 7. Core and coil arrangements of a shell-type high-reactance transformer derived from Fig. 6(B)

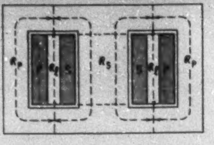


Fig. 8. Core and coil arrangements of a shell-type low-reactance transformer

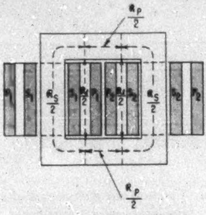

$\frac{\mathcal{R}_P}{2}$
P_1
S_1
$\frac{\mathcal{R}_S}{2}$
S_1
$\frac{\mathcal{R}_l}{2}$
P_1
P_2
$\frac{\mathcal{R}_l}{2}$
S_2
$\frac{\mathcal{R}_S}{2}$
S_2
P_2
$\frac{\mathcal{R}_P}{2}$

Fig. 9. Core and coil arrangements of a core-type low-reactance transformer

inductance $L_{P,m}$. The reluctance $\mathcal{R}_S$ and the turns of the secondary coil similarly determine the value of secondary magnetizing inductance, $L_{S,m}$. (In Figs. 5–7, an air gap has been shown in the secondary-core portions merely to indicate one factor which can cause $\mathcal{R}_P$ and $\mathcal{R}_S$ to be different.) The reluctance $\mathcal{R}_l$ and the turns of the secondary coil are factors which determine the leakage inductance referred to the secondary, L_l.

If the lengths, areas, and core materials constituting the primary and secondary magnetic circuits are all identical, then the inductances $n^2L_{P,m}$ and $L_{S,m}$ will be equal for equal flux densities. However, if identical conditions do not exist, such as through inequalities of core areas, flux densities, lengths of flux paths, or unequal air gaps in the primary- and secondary-core portions, then the inductances $n^2L_{P,m}$ and $L_{S,m}$ may not be equal.

With these relationships between the physical portions of the transformer and the inductive elements of the π-equivalent clearly established for a relatively simple form of high-leakage-reactance transformer, a series of changes in coil and core configurations will serve to demonstrate that similar direct relationships exist even in a conventional low-reactance transformer.

Fig. 6 shows two coil configurations on the same core configuration which are magnetically equivalent in respect to core fluxes and reluctances; i.e., the differences due to leakage fluxes external to the core structure are disregarded. Fig. 6(A) illustrates a well-known type of high-reactance transformer. High-reactance transformers illustrated by Fig. 6(B) are not so well known. This latter type is employed when it is desirable to keep the air-path leakage flux between primary and secondary to a minimum. Note that although portions of the primary coil conductors which are external to the core window are in different positions in Fig. 6(B) than in Fig. 6(A), they have the same portions of their respective core structures associated with them. From this fact it can be concluded that it is only the relative positions of the primary and secondary coil conductors in the core window that determine the portions of the core with which they are associated.

The configurations of Figs. 5 and 6 are commonly referred to as "core-type" high-reactance transformers. Fig. 7 is a "shell-type" high-reactance transformer derived from Fig. 6(B). It is not a practical configuration, the usual shell-type high-reactance transformer being of the type derived from Fig. 5. However, Fig. 7 better serves the transition to the conventional low-reactance (close-coupled) transformer core and coil arrangement shown by Fig. 8. The leakage flux in low-reactance transformers is entirely in air as shown in this figure. However, the area of leakage flux path is not confined to the window as illustrated, but also includes the portions of the coil external to the window. This figure also shows that in the shell-type transformer with a single primary and single secondary, the actual length of core associated with the primary is substantially the same as that associated with the secondary.

Applying the same principles of relating the windings to specific core portions of the conventional core-type low-reactance transformer produces the results shown in Fig. 9. This is of particular interest as it shows that for the particular relative positions of the primary and secondary coils illustrated, the secondary-core portions are much longer than the primary-core portions. In normal designs of this coil arrangement, the secondary magnetizing inductance will therefore be less than the primary magnetizing inductance as reflected to the secondary. (Interchanging primary and secondary coil positions will, of course, cause the converse of this.) A similar condition will be found to exist in a shell-type transformer when the secondary consists of two coils of equal turns with the primary coil between them.

The foregoing examples for a variety of core and coil configurations demonstrate that it is possible to relate the magnetic characteristics of appropriate structural parts of a particular transformer to the inductive elements of a π-equivalent circuit for that transformer. The validity of this division of the core into primary- and secondary-core portions is further substantiated by experimental data which follow.

Methods for Determining Equivalent Circuit Parameters

The leakage inductance of a transformer under test is measured in the conventional manner as shown in Fig. 10. This particular arrangement of instruments is required when measuring high-reactance transformers in order to provide measurements which are insensitive to the effects of magnetizing currents required by core portions other than those associated with the leakage flux.

Measurement of the two inductances corresponding to the primary and secondary magnetizing inductances requires

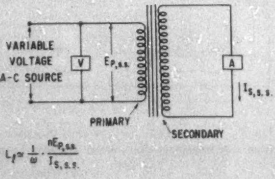
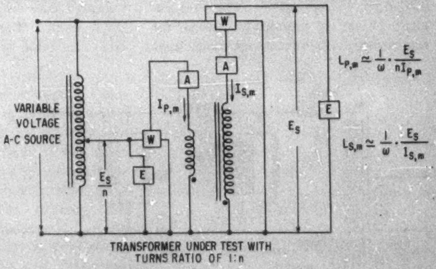

VARIABLE VOLTAGE A-C SOURCE
V
$E_{P,s.s.}$
A
$I_{S,s.s.}$
PRIMARY
SECONDARY
$L_l \simeq \frac{1}{\omega} \cdot \frac{nE_{P,s.s.}}{I_{S,s.s.}}$

Fig. 10 (left). Circuit for determining leakage inductance

Fig. 11 (right). Circuit for determining primary and secondary magnetizing characteristics


W
A
A
$I_{P,m}$
$I_{S,m}$
E
VARIABLE VOLTAGE A-C SOURCE
W
E
E_S
$\frac{E_S}{n}$
$L_{P,m} \simeq \frac{1}{\omega} \cdot \frac{E_S}{nI_{P,m}}$
$L_{S,m} \simeq \frac{1}{\omega} \cdot \frac{E_S}{I_{S,m}}$
TRANSFORMER UNDER TEST WITH TURNS RATIO OF 1:n

a technique which is radically different from that usually employed when the magnetization characteristics of the entire core are determined by an open-circuit test. Fig. 11 is a schematic circuit diagram of a test circuit which provides data for determining the characteristics of $L_{P,m}$ and $L_{S,m}$. Both the primary and secondary are excited simultaneously with in-phase voltages E_S/n and E_S, respectively. When the proper transformer terminals are so excited, the leakage flux will be zero; therefore, the current flowing in the primary winding will be that required to excite the core portion associated with the primary, and the current flowing in the secondary will be that required to excite the core portion associated with the secondary. (These separate excitation currents do not necessarily flow under normal operating conditions.) If the core losses are small as compared with the reactive components of exciting current, the relationships given in Fig. 11 provide reasonably accurate values to use for the equivalent circuit. If the core losses are not negligible, then the wattmeter readings provide information for use in determining the values of resistances to be used as shunt elements in the equivalent circuit of Fig. 3 and for more accurately determining the values of the magnetizing inductance elements for the equivalent circuit.

Core-type transformers with primary and secondary coils on each leg have a core and coil configuration, the physical relationship of which to the π-equivalent is the least obvious of all of the simpler configurations. Also, as pointed out earlier, normal designs of this type call for an equivalent circuit whose shunt or magnetizing inductance elements are not equal. For these reasons, a core-type transformer was chosen to demonstrate how the test methods outlined are effective in determining the separate magnetizing characteristics of the three inductive components of the transformer which correspond with the three inductive elements of the π-equivalent circuit.

Two high-reactance 1-to-1-ratio core-type transformers were constructed, the respective coils of which were identical. They were similar in physical appearance to that depicted by Fig. 9 with the exception that magnetic shunts were placed between the primaries and secondaries in the position identified in Fig. 9 by the symbol $\mathcal{R}_l/2$, to provide high leakage inductance. The cores of these two units were substantially identical in over-all dimensions. However, in unit *A* an air gap was placed in the center of each yoke to increase the reluctance of the primary-core portions, whereas in unit *B* an air gap was placed in the center of each leg to increase the reluctance of the secondary-core portions.

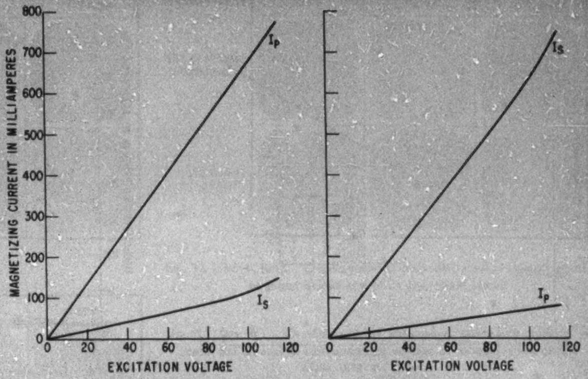


Fig. 12. Excitation volt-ampere characteristics

A—Test unit A, air gaps in yokes (parts of core associated with primary)
B—Test unit B, air gaps in legs, under coils (parts of core associated with secondary

The results of short-circuit impedance tests showed that the leakage inductance characteristics of the two units are approximately the same.

The results of primary and secondary magnetizing impedance tests of the two units are shown in Fig. 12. These data clearly show the effect of the position of the air gap with respect to the primary- and secondary-winding-core portions on the magnetization characteristics of the primary and of the secondary. In unit *A*, with the air gap in the yoke, the exciting current of the outer coil, which has been arbitrarily labelled the primary, is much higher than that of the secondary or inner coil. But in unit *B*, with the air gaps in the leg portions, the primary exciting current is very low while the secondary exciting current is high. Note also that the lower curve of unit *B*, which is a measure of the exciting requirements of the yoke-core material, shows values for corresponding voltages which are lower than those of the lower curve for unit *A*, which corresponds to the exciting current requirements of the core material of the leg sections. This relationship is consistent with the fact that the yoke-core portions are shorter than the leg-core portions; hence at a given flux density they should require lower volt-amperes of excitation than are required by the leg-core portions.

From the test results displayed by Fig. 12, it seems reasonable to draw the following conclusions:

1. The test method described in connection with Fig. 11 does provide a means for measuring the primary and secondary magnetization characteristics of a transformer.
2. It is possible, even in rather complex coil configurations, to relate certain core portions and their excitation requirements to appropriate coil windings.
3. The primary and secondary magnetizing impedances in certain transformer coil and core configurations are inherently unequal.
4. A direct relationship can be established between the magnetic reluctances of certain portions of the transformer-core structure and the inductive elements of its π-equivalent circuit.

Consideration of Saturation Effects

In some transformer applications the secondary-core portion may be operated in the saturation region at the same time that the primary is operating at somewhat lower flux densities. Such a mode of operation may arise from two effects which can be present either singly or in combination: (1) reduced cross section of the secondary-core portion compared with that of the primary as in a peaking transformer, and (2) an effective leading-power-factor load on the secondary, such as is present in regulator-type transformers and ballasts. Since the π-equivalent circuit separates the primary magnetizing inductance from the secondary magnetizing inductance, it provides a means for including the effects of these differences in the equivalent circuit. This can be done by having the shunt elements of the π-equivalent circuit which represent the primary and secondary exhibit the same degree of nonlinearity as is exhibited by

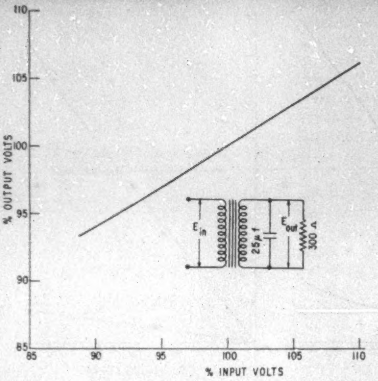


Fig. 13. Characteristics of modified unit A operating as an elementary voltage stabilizer

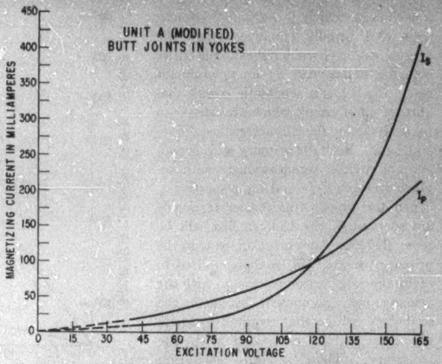


Fig. 14. Excitation volt-ampere characteristics of modified unit A over voltage ranges pertinent to its operation as a voltage stabilizer

the transformer of which it is designed to be an equivalent. The test method described in connection with Fig. 11 provides a means for determining the respective volt-ampere characteristics of the primary and of the secondary. It is required, of course, that the input voltages under test be carried high enough to include the highest voltage known to appear across a given winding under operating conditions. (When the secondary cross sections of the core is drastically reduced, as in a peaking transformer, it may not be possible to carry the test so as to bring the primary voltage to full excitation value. However, in this case, such a restriction is not serious since the primary magnetizing inductance plays no part in determining the output waveform characteristic, and therefore its effect can be neglected.)

The transformer identified as unit *A* exhibits regulating properties if its secondary is paralleled by a 25-microfarad capacitor and a resistance load of 300 ohms. Under these conditions, the secondary voltage was 156 volts with a primary input voltage of 118 volts. Fig. 13 shows the regulation curve of secondary voltage versus primary voltage over a range of 105–130 volts on the primary. (The air gaps in the yokes were reduced to butt joints in order to reduce the primary exciting current for these tests.) Fig. 14 shows the results of impedance tests on this unit to determine the proper volt-ampere characteristics required of the corresponding shunt impedances to use in a π-equivalent circuit for this transformer under these operating conditions.

Physical Equivalent Circuits and Their Operation

Owing to the close relationship between the π-equivalent circuit element and the transformer of which it is an equivalent, it is possible to construct physical equivalent circuits which perform in the same manner as their equivalent transformers. This can be done even though saturation effects are present in certain of the elements under operating conditions. The test results which follow provide graphic proof that this is possible. They also provide a convincing demonstration of the validity of this π-equivalent circuit.

Fig. 15 shows a schematic diagram of the equivalent circuit of the type of device selected to demonstrate this close tie between a high-reactance transformer and its equivalent circuit. The device is a ballast for arc-discharge types of lamps. This ballast consists of a high-reactance transformer with a capacitor in series with the transformer secondary and the arc-discharge lamp load.

The transformer parameters for three ballasts of widely different lamp loads were determined in accordance with the methods described in this paper. From the information obtained, coupled with the primary and secondary currents and voltages under both open-circuit and operating conditions, inductors and transformers for π-equivalent circuits were designed and built. The inductor elements for the equivalent circuits were then adjusted so that their volt-ampere characteristics corresponded very closely to the parameters of the particular transformer which they were designed to represent. The comparative results of elec-

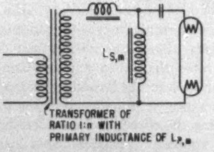
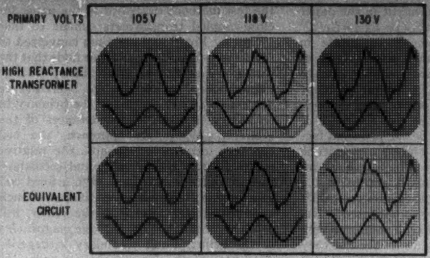
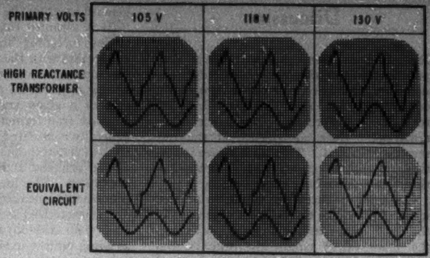


Fig. 15. Simplified equivalent circuit of a high-reactance transformer with a series-connected capacitor and arc-discharge-type lamp as load

Table I. Results of Electrical Measurements

Ballast No.	Line Volts	Secondary Current, Amperes		Lamp Input, Watts	
		Ballast	Equivalent	Ballast	Equivalent
	106	10.7	10.8	1,140	1,160
1	118	10.9	11.1	1,150	1,180
	130	11.1	11.4	1,170	1,200
	106	1.99	2.02	244	246
2	118	2.06	2.08	249	252
	130	2.11	2.14	254	256
	106	0.383	0.394	66.0	65.8
3	118	0.396	0.404	67.9	66.8
	130	0.401	0.406	68.5	67.8





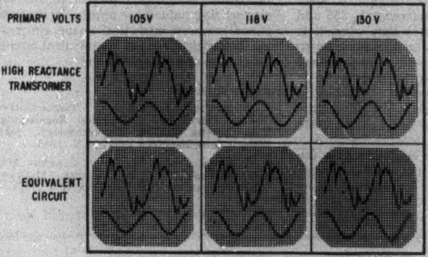


Fig. 16 (above, left). Oscillograms of secondary current (lamp current) of actual ballast and of its equivalent circuit

Fig. 17 (above). Oscillograms of transformer secondary voltage of actual ballast and of corresponding voltage of its equivalent circuit

Fig. 18 (left). Oscillograms of voltage induced in test coils around magnetic shunts of actual ballast transformer and of voltage across series inductor of its equivalent circuit

trical measurements on these transformer ballasts and their respective equivalent circuits when operating into their rated lamp loads are shown in Table I. Considering the fact that nonlinear effects were present in the secondary-core portions of all of these units, and that the lamp loads also exhibit nonlinear volt-ampere characteristics, the close correspondence between the operating characteristics of the physical equivalent circuit and its real counterpart shown by Table I is quite good.

The results of oscillographic tests on one of the ballasts and its equivalent circuit are included in Figs. 16–18. These show that the differences between corresponding waveshapes of the ballast and the equivalent circuit are extremely small. Fig. 18 is particularly interesting as it demonstrates the accuracy with which the series linear inductor duplicates the effect of the magnetic shunts in the actual high-reactance transformer.

Conclusions

The seldom-used π-equivalent circuit has been shown to be capable of accurately representing a transformer in which nonlinear effects are present. Also, the inductive elements of this equivalent circuit have been shown to be directly relatable to the magnetic characteristics of certain portions of the transformer core. In addition, a physical π-equivalent circuit can be constructed which will operate in almost precisely the same manner as the transformer of which it is an equivalent. In view of these several advantages, the π-equivalent circuit should be employed in analyzing the operation of circuits which include a transformer as a circuit element when the transformer is a high-reactance type and/or when saturation effects must be taken into account in the analysis.

When saturation effects are employed in a high-reactance transformer to provide a desired operating characteristic, the π-equivalent circuit, with its accurate portrayal of a transformer, provides a means for expediting the development of the required transformer. Since the parameters of a developmental physical equivalent circuit can be varied independently, the transformer designer can vary the parameters of the transformer equivalent circuit along with other circuit parameters, until the complete circuit provides the desired operating characteristics. The transformer design procedure then becomes one of attaining a core and coil configuration with magnetization characteristics which correspond to the pertinent elements of the developmental equivalent circuit. The close relaships between the transformer-core portions and the corresponding elements of the equivalent circuit enable the designer to use the pertinent information supplied by the developmental equivalent circuit, coupled with the basic core material's magnetic properties, to determine the proper physical shapes and dimensions of various parts of the transformer-core structure which will provide the desired magnetization characteristics. Of course, this part of the design procedure may show that the magnetic properties of the magnetic material the designer wishes to employ impose restrictions which prohibit the attainment of the desired result. However, it is much better to find it out through the use of a developmental equivalent circuit and a bit of calculation than to have to draw the same conclusion after a time-consuming series of tests on a number of successive developmental transformer designs.

It is hoped that this paper, coupled with the increased interest in nonlinear magnetics, will create such an interest in the π-equivalent circuit for transformers that it will soon become generally recognized as the proper tool to use in the analysis and design of transformers in which nonlinear effects are present.

Reference

1. Transformer Engineering, L. F. Blume et al. John Wiley & Sons, Inc., New York, N. Y., first edition, 1938, pp. 64–68.

Discussion

Reuben Lee (Westinghouse Electric Corporation, Baltimore, Md.): Mr. Lord has demonstrated that the π-circuit is a better equivalent than the T-circuit for high-reactance transformers. This paper shows remarkably close equivalence in the test results of Table I. This equivalence is particularly appropriate for the several applications which have been chosen.

One minor point which might be omitted without prejudice to the main argument of the paper concerns the assumption that in the T-equivalent circuit the primary and secondary leakage inductances are equal when referred to the same winding. This is only true for especially designed coils, and it is by no means true in the general case. Also, it is not necessary in the equivalent T-circuit, where it is applicable, to regard the leakage inductances of the two windings as equal.

H. W. Lord: I agree with Mr. Lee's statement that in the T-equivalent circuit the primary and secondary leakage inductances are not required to be equal nor are they equal in the general case. I did not intend to imply that such equality existed, in the general case, by the statement in the paper "--it is assumed that $n^2L_{P,l}=L_{S,l}=L_l/2$."

The foregoing statement was made in connection with the determination, by conventional test measurements, of the values to use for $L_{P,l}$ and $L_{S,l}$ in the T-equivalent circuit of a given transformer. In a close-coupled transformer, the coil geometry is generally considered to be the principal factor which determines the relative magnitude of the primary and secondary leakage inductances. The methods described in the literature for measuring their relative values are characterized as very complicated (reference 1, pp. 354–55) and such measurements as being subject to large errors because they require a measurement of the magnetizing inductance to an accuracy of about 0.01% in order to secure a 5% accuracy in the ratio of their magnetidues (reference 1, p. 454). Therefore, textbooks (reference 1, 355 and reference 2, p. 188) state that when it is necessary to make any assumptions as to their relative magnitudes, it is often assumed as an approximation that they are equal. It was against this background of practice in relation to the T-equivalent circuit that the statement referred to by Mr. Lee was made in the paper.

Considerations of the equivalency of the T- and π-circuits show that if the π-equivalent-circuit parameters are determined by the test methods described in the paper for close-coupled transformers similar to that shown by Fig. 9 and are then converted to their equivalent T-circuits, one finds that the ratios of the primary to secondary leakage inductances of such equivalent T-circuits are determined by the ratios of the primary to secondary magnetizing inductances of the π-circuit. With this approach, the coil geometry does not determine the relative magnitudes of the primary and secondary leakage inductances since, with a fixed coil geometry, a change in core reluctances such as in test units *A* and *B* of the paper, will produce a change in the relative magnitudes of the primary and secondary leakage inductances of the T-equivalent circuit. It was because of this difference between T-equivalent circuits arrived at by the conventional coil geometry approach and by the magnetizing inductance approach, that the statements were made in the paper concerning the deficiency of the T-equivalent circuit for close-coupled transformers.

REFERENCES

1. MAGNETIC CIRCUITS AND TRANSFORMERS (book), Department of Electrical Engineering, Massachusetts Institute of Technology. John Wiley & Sons, Inc., New York, N. Y., 1943.

2. PRINCIPLES OF ALTERNATING CURRENT MACHINERY (book), Ralph R. Lawrence. McGraw-Hill Book Company, Inc., New York, N. Y., 1921.

SPUD, A Stored-Program Universal Demonstrator for Computer Training

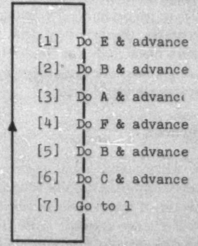
M. RASPANTI
NONMEMBER AIEE

CONSIDERABLE difficulty is usually encountered by the novice in understanding how stored-program information-processing systems operate, and how they can be programed. While conventional teaching tools are definitely inadequate in this case, the use of a commercial computer for demonstrations does not provide an ideal solution either, because the machine tends to overwhelm the novice with its speed and complexity. A satisfactory solution has been found in SPUD (Stored-Program Universal Demonstrator), a stored-program relay machine designed specifically as a teaching aid, complicated enough to stimulate the novice but still easy to grasp in its entirety.

Wired and Stored Logic

The basic principles of information-processing machines can be reviewed very conveniently by means of an analogy. Assume that we want to train a native from some uncivilized region. To enable him to perform some specific task, we must somehow show him exactly what to do under various conditions. At the end of the training, we will have established in the native's brain the mental paths that he needs to handle the job on his own. In a somewhat similar fashion, we can design a machine to perform a specific task by connecting a number of switching devices in such a way as to create the necessary paths of mechanized reasoning. In this case, we say that the machine uses only "wired logic" because the intelligence to perform a specific job resides in the wiring of the machine. This arrangement has an inherent lack of flexibility: an extensive rewiring or even a major redesign may be required to modify the task to be performed by the machine.

The native of the analogy suffers from a similar disadvantage: he must be trained again whenever a new task is assigned to him. This difficulty can be eliminated if we can establish some language with which to communicate with our native. Instead of giving him a specialized training for each specific job, we will train him once and for all to understand and execute any one of a set of instructions or orders. We can then



```
[1] Do E & advance
[2] Do B & advance
[3] Do A & advance
[4] Do F & advance
[5] Do B & advance
[6] Do C & advance
[7] Go to 1
```

Fig. 1. A single-branch program

Paper 59-665, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Baltimore, Md., May 19–21, 1959. Manuscript submitted October 31, 1958; made available for printing March 19, 1959.

M. RASPANTI is with the Bell Telephone Laboratories, Inc., Whippany, N. J.

The author wishes to acknowledge the contribution of J. B. Worth, who provided the mechanical design of the program storage unit and housing, and also the several interesting discussions and suggestions by A. H. Budlong, J. A. O'Brien, and W. Keister, all of Bell Telephone Laboratories.

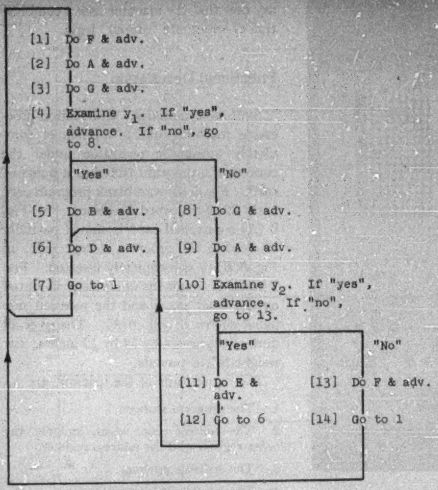
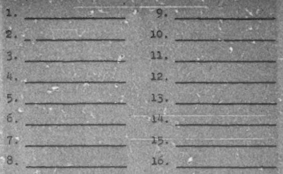
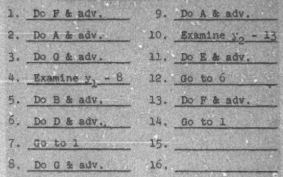


Fig. 2 (left). A multibranch program



1. ____________ 9. ____________
2. ____________ 10. ____________
3. ____________ 11. ____________
4. ____________ 12. ____________
5. ____________ 13. ____________
6. ____________ 14. ____________
7. ____________ 15. ____________
8. ____________ 16. ____________

Fig. 3 (right). Blank program form



1. Do F & adv. 9. Do A & adv.
2. Do A & adv. 10. Examine y_2 - 13
3. Do G & adv. 11. Do E & adv.
4. Examine y_1 - 8 12. Go to 6
5. Do B & adv. 13. Do F & adv.
6. Do D & adv. 14. Go to 1
7. Go to 1 15. ____________
8. Do G & adv. 16. ____________

Fig. 4 (right). Program form filled out for the example of Fig. 2

enable him to perform a specific job by writing an appropriate sequence of instructions in the language that he understands. One such sequence of instructions, or program, is shown in Fig. 1. In the DO type of orders, the letters *E*, *B*, *A*, etc., are used to designate elementary actions to be performed by the native, who is also instructed to advance to the next order. At the end of the program, a GO type of order instructs the native to transfer to some specified instruction. Observe that, in any particular program, some of the orders in the native's repertoire may not be used at all, while others may be used quite frequently.

The type of program shown in Fig. 1 can prescribe only a single line of action. A very flexible type of program can be written if we train our native to examine or "read" any one of a number of situations $y_1, y_2, \ldots, y_t$ and to determine which of n possible cases exists at the time. We can then give the native very definite instructions for each of the n possible cases. If each situation results in either a "yes" or a "no" case, we can write a program of the type shown in Fig. 2.

Regardless of the pattern in which the program branches out, the instructions can be submitted to the native in a very orderly fashion by means of a blank program form of the type shown in Fig. 3. A complete instruction or "word" can be written in each blank line or "word location," which is uniquely identified by a number or "address." Fig. 4 shows the program of Fig. 2 in the form in which it would be submitted to the native.

Merely by following to the letter the instructions of the program, the native is capable of performing some specific task under the various conditions that may arise. The intelligence to perform the specific task is not any longer in the native's brain; it resides, instead, in the form in which the program has been written. Without any additional training, the same native can perform an entirely different task if he is supplied with the appropriate program. In fact, the native can perform any task for which a program can be written in his language using no more than 16 "words."

In a similar fashion, a program-controlled machine uses general-purpose wired logic to interpret and execute any one of a number of instructions. For the specific tasks to be performed, programs are written in machine language and recorded in some storage or memory unit, from which they are fed to the implementing circuitry of the machine. We

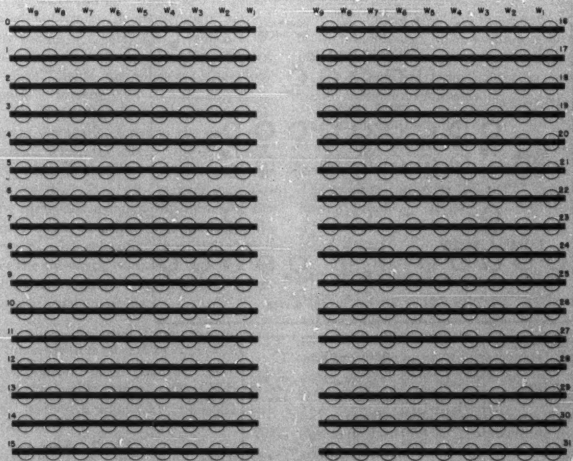

w_9 w_8 w_7 w_6 w_5 w_4 w_3 w_2 w_1
w_9 w_8 w_7 w_6 w_5 w_4 w_3 w_2 w_1
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Fig. 5. Blank program card

say now that the machine uses a combination of wired and "stored" logic.

Functional Description

Spud is a miniature information-processing machine that is capable of completely automatic operation under the control of a program stored in a punched card. Fig. 5 shows a blank program card with 32 9-bit word locations. In Fig. 6 (A) a punched card is shown partially inserted in the program storage unit; in Fig. 6(B), it is completely inserted. Fig. 7 labels in detail the equipment mounted on the front panel and the recessed side panel shown in Fig. 6(A). The over-all dimensions are 24 by 34 by 12 inches; the weight is 110 pounds.

The major units of the machine are

1. The program storage.
2. The control unit, which includes the order register and the address register.
3. The erasable memory.
4. The input unit.
5. The output unit.

Only the program storage will be described in some detail; the remaining units will be given a brief functional description

The Program Storage

As shown in Fig. 8, the program storage consists of two matrices, each with 16 rows and 9 columns; the rows correspond to word locations, the columns to bits within a word. At each cross-point, a contact element tends to connect the corresponding row and column. Whether a connection is actually established depends on whether a corresponding hole has been punched on the program card, which is inserted between the contact elements and the columns.

At any one time, ground is applied to only one row via a selecting circuit.[1] This ground is allowed to appear only on those columns for which, along the grounded row, holes have been punched on the program card; e.g., a ground on row 15 will appear on column W_1 but not on W_9. The diodes are used to prevent ground backups. For instance, when row 15 is grounded, if no diodes were provided, the ground would be allowed to reach column W_9 over row 0, even though there is no hole at the cross-point of row 15 and column W_9. Thus, grounding a row reads out the corresponding word by converting a combination of punched holes on that row into a combination of ground signals on nine leads to the control unit.

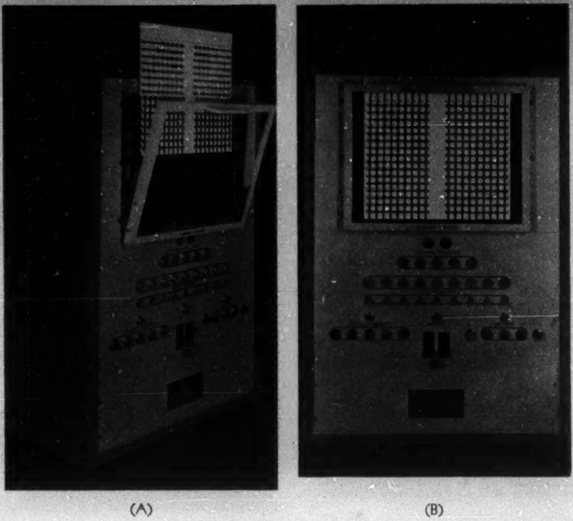

(A) (B)

Fig. 6. Front view of SPUD

A—Program card partially inserted
B—Program card completely inserted

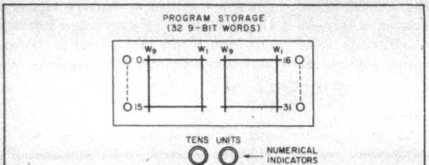
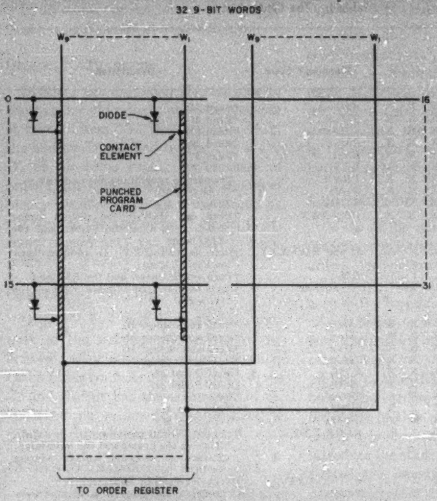

PROGRAM STORAGE
(32 9-BIT WORDS)
W_9
W_1
W_9
W_1
0
16
15
31
TENS
UNITS
NUMERICAL INDICATORS
ADDRESS REGISTER
A_5
A_4
A_3
A_2
A_1
ORDER REGISTER
W_9
W_8
W_7
W_6
W_5
W_4
W_3
W_2
W_1
"MANUAL WORD" SW.'S
RM
SI
RDO
ERASABLE MEMORY
INPUTS
OUTPUTS
M_5
M_4
M_3
M_2
M_1
"1"
"0"
P
Q
POWER
Z_3
Z_2
Z_1
CLOCK
RELEASE PUSH BUTTON
FRONT PANEL
FUSES
MODE SELECTOR SWITCH
AW-MK
MW-MK
AW-AK
RECESSED SIDE PANEL
Q
P
MANUAL CLOCK
POWER SWITCH
ON
OFF
SPARE
RECEPTACLES

Fig. 7. Front and side panel arrangement



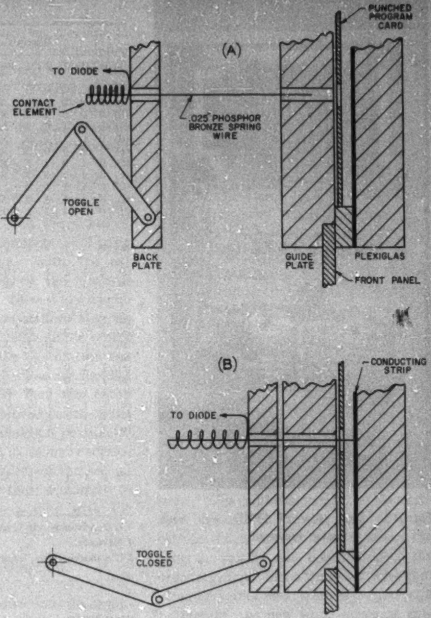


Fig. 8 (above). Electrical schematic of program storage

Fig. 9 (right). Mechanical schematic of program storage

Some mechanical details of the program storage are given in Fig. 9. Each contact element consists of a coil spring which continues into a straight wire along the axis of the coil. The 288 contact elements are mounted on a back plate, which can be moved toward or away from the front panel by means of a toggle controlled by the frame shown in Fig. 6(A) with the engraving "program storage." When the frame is in the position of Fig. 6(A), the back plate is furthest away from the front panel. The straight portions of the contact elements move through holes in a stationary guide plate mounted on the front panel. Eighteen conducting vertical strips are embedded in the back of a Plexiglas plate mounted on the front panel. When the toggle is open, Fig. 9(A), the straight wires retract into the guide holes; a program card can be dropped now in the slot between the guide plate and the Plexiglas plate. When the toggle is closed, Fig. 9(B), the straight wires make contact with the corresponding conducting strips wherever holes have been punched on the program card.

The Control Unit

The function of this unit is to interpret and execute the orders that are read out of the program storage, one at a time in synchronism with a clock pulse. The clock pulse may be generated manually by means of a push button plugged into a receptacle on the recessed side panel (Figs. 6(A) and 7). It may also be generated automatically about every second by means of a small motor, which actuates a set of contacts. The lamp labeled "clock" lights up whenever the clock pulse is present.

The order that is read out of the program storage is stored in the order register. While the clock pulse is present, the order is displayed on the nine lamps W_1 to W_9. The order register may also be controlled manually by means of the nine toggle switches under the lamps W_1 to W_9. By selecting either the manual or the automatic control for the clock and the order register, three modes of operation are made possible:

1. Automatic word (from the program storage) with manual clock (*AW-MK*).
2. Manual word (set on the toggle switches) with manual clock (*MW-MK*).
3. Automatic word with automatic clock (*AW-AK*).

The mode of operation is selected by means of the rotary switch mounted on the recessed side panel.

An important function of the control unit is to determine the address of the next order to be read out of the program storage. This address is stored in the address register and is transmitted to the program storage after the disappearance of the clock pulse. The address is displayed in binary form on the five lamps A_1 to A_5, and in decimal form on two Nixie* numerical indicators.

The Erasable Memory

This unit consists of five elements monitored by the five lamps M_1 to M_5. The machine can write "0" or "1" in any memory element; it can also read the information stored in any memory element.

The Input Unit

Information is fed into the machine via two input channels which are normally controlled by the keys P and Q. In addition to the up or "1" position and the down or "0" position, each key has a middle position to transfer the control of the corresponding channel to a remote switch plugged into two receptacles on the recessed side panel.

---

* Registered trade-mark of the Borroughs Corporation.

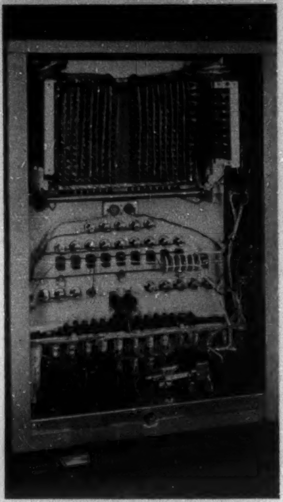


Fig. 10. Rear view of SPUD, with back cover removed

Table I. The Order List

	Binary Code									Mnemonic Code	Description
	W_9	W_8	W_7	W_6	W_5	W_4	W_3	W_2	W_1		
(a)	0	0	0	0	X	X	X	X	X	RM_1-x	*Read M_1
(b)	0	0	0	1	X	X	X	X	X	RM_2-x	*Read M_2
(c)	0	0	1	0	X	X	X	X	X	RM_3-x	*Read M_3
(d)	0	0	1	1	X	X	X	X	X	RM_4-x	*Read M_4
(e)	0	1	0	0	X	X	X	X	X	RP-x	*Read P
(f)	0	1	0	1	X	X	X	X	X	RQ-x	*Read Q
(g)	0	1	1	0	X	X	X	X	X	TRA-x	Transfer to program storage address x
(h)	0	1	1	1	0	1	X	X	X	W_1M_x	†Write 1 in M_x
(i)	0	1	1	1	0	0	X	X	X	WOM_x	†Write 0 in M_x
(j)	0	1	1	1	1	0	X	X	X	RDO-XXX	†Read out $W_3W_2W_1$ on output lamps $Z_3Z_2Z_1$. Turn on the buzzer and the RDO lamp
(k)	0	1	1	1	1	1	X	X	X	SI and RDO-XXX	†Same as RDO-XXX. In addition, turn on the SI lamp
(l)	0	1	1	1	1	0	1	1	1	RM	†Turn on the buzzer and the RM lamp
(m)	0	1	1	1	1	1	1	1	1	SI and RM	†Same as RM. In addition, turn on the SI lamp
(n)	1	0	0	0	X	X	X	X	X	CM_1-x	*Read and change M_1
(o)	1	0	0	1	X	X	X	X	X	CM_2-x	*Read and change M_2
(p)	1	0	1	0	X	X	X	X	X	CM_3-x	*Read and change M_3
(q)	1	0	1	1	X	X	X	X	X	CM_4-x	*Read and change M_4
(r)	1	1	0	0	X	X	X	X	X	RQM_1-x	‡Read Q and M_1
(s)	1	1	0	1	X	X	X	X	X	RPM_2-x	‡Read P and M_2
(t)	1	1	1	0	X	X	X	X	X	RM_5-x	*Read M_5
(u)	1	1	1	1	0	0	X	X	X	WPM_x	†Write P in M_x
(v)	1	1	1	1	0	1	X	X	X	WQM_x	†Write Q in M_x
(w)	1	1	1	1	1	0	X	X	0	$RDOR$-XXO	Read out W_3W_2O on output lamps $Z_3Z_2Z_1$. Turn on the buzzer and the RDO lamp. Return to program storage address 00000
(y)	1	1	1	1	1	X	X	X	1	$W3WM$-XXX	†Write $W_4W_3W_2$ in $M_4M_3M_2$
(z)	1	1	1	1	1	1	0	0	0	TRM	Transfer to the program storage address in the erasable memory

* If 0, advance. If 1, transfer to program storage address x.

† Advance.

‡ If a match exists, advance. If a mismatch exists, transfer to x.

THE OUTPUT UNIT

This unit comprises a buzzer and the lamps Z_1, Z_2, Z_3, RDO, SI, and RM. Normally, the machine gives out information as 3-bit words displayed on the lamps Z_1, Z_2, and Z_3; all the combinations are possible except 111. When the machine gives an output, the following actions take place:

1. The "read the output" lamp RDO lights up and the buzzer sounds to attract the attention of the operator.
2. The machine locks itself up making any subsequent clock pulses ineffective.

At the time the machine gives an output, it may also request the operator to set a new input on P and Q by lighting the "set the input" lamp SI. The operator acknowledges the output by pressing the release push button. This action extinguishes the lamps Z_1, Z_2, Z_3, SI, and RDO, shuts off the buzzer, and unlocks the machine.

The machine may also give out information by utilizing part or all of the erasable memory as an output unit. The actions that take place in this case are:

1. The "read the memory" lamp RM lights up and the buzzer sounds.
2. The machine locks itself up.

As for the previous case, the machine may request at the same time a new input by lighting the lamp SI. The operator acknowledges the output again by pressing the release push button. This action shuts off the buzzer, extinguishes the lamps RM and SI, and unlocks the machine. The state of the erasable memory is not affected in any way.

THE INTERNAL MECHANICAL ARRANGEMENT

Fig. 10 shows the arrangement of the equipment inside the machine. Twenty-nine relays are mounted behind the lower part of the front panel. The power supply and the motor for the automatic clock are mounted on the chassis at the bottom. All the connections to this chassis may be easily unplugged, and the chassis may be removed to give access to the relay terminals. The upper part of Fig. 10 shows the coil springs mounted on the back plate of the program storage. The nine diodes for each word location are mounted on a printed card of the type shown near the bottom of the machine. The diode cards are plugged into the connectors shown on either side of the back plate of the program storage. Access to the connector terminals is gained by removing the printed cards and rotating the hinged connector assemblies by 90 degrees toward the program storage.

The Order List

Table I lists all the orders that the machine can interpret and execute. For each order, the table gives the 9-bit binary code, an abbreviation or mnemonic code, and a brief description. The orders may be classified as follows.

ORDERS TO READ

Orders *a*, *b*, *c*, *d*, *e*, *f*, and *t* instruct the machine to read a specified memory element or input channel. If a "0" is read, the machine shall advance to the next address in the program storage. If a "1" is read, the machine shall transfer to the program storage address specified by the five bits W_5 (most significant) to W_1 (least significant).

ORDERS TO WRITE

Orders *h* and *i* instruct the machine to write "1" or "0," respectively, in the memory element specified by the bits W_3 to W_1.

Orders *u* and *v* instruct the machine to "copy" the state of input channel P or Q, respectively, into the memory element specified by the bits W_3 to W_1.

Order *y* instructs the machine to "copy" the bits W_4, W_3, and W_2 into the memory elements M_4, M_3, and M_2, respectively.

In all cases, the machine shall advance to the next program storage address.

ORDERS TO TRANSFER

Orders *g* and *z* instruct the machine to transfer unconditionally to some program storage address. In the case of order *g*, the address is specified by the bits W_5 to W_1 of the order itself. In the case of order *z*, the address is given by whatever binary number happens to be stored at the time in the five memory elements M_5 (most significant) to M_1 (least significant).

ORDERS TO READ OUT

Orders *j*, *k*, *l*, and *m* control the outputs of the machine as previously described; they all result in an advance. Order *w* can control only the output lamps Z_3 and Z_2 but has the advantage of resulting in an automatic transfer back to program storage address 00000.

ORDERS TO CHANGE

Orders *n*, *o*, *p*, and *q* instruct the machine to read and then change memory elements M_1, M_2, M_3, and M_4 respectively. If a "0" is read (before the change), the machine shall advance to the next program storage address. If a "1" is read, the machine shall transfer to the address specified by the five bits W_5 to W_1.

ORDERS TO MATCH

Orders *r* and *s* instruct the machine to read simultaneously input Q and memory element M_1, or input P and memory element M_2, respectively. If a match exists, the machine shall advance; if a mismatch exists, the machine shall transfer to the program storage address specified by the five bits W_5 to W_1.

8-BIT AND 9-BIT ORDERS

It should be noted that the group of orders *a* to *m* forms a complete set of basic orders for programing. With the exception of order *t* (RM_5–x), the result of each of the remaining orders *n* to *z* can be accomplished by a combination of orders from the first group. Since all the orders *a* to *m* have "0" in the W_9 bit, we can pretend at first that the machine handles only 8-bit orders and that the entire order list of the machine consists of the orders *a* to *m*. This makes it possible to introduce the students to the more sophisticated time-saving orders of the second group only after they have acquired sufficient facility in programing with the simpler orders of the first group.

The General Block Diagram

In the preceding two sections the organization of the machine and its "language" were discussed. This information can be co-ordinated now by referring to the block diagram of Fig. 11. Although relays are used in the machine, gate symbols are employed in the block diagram as convenient functional blocks.

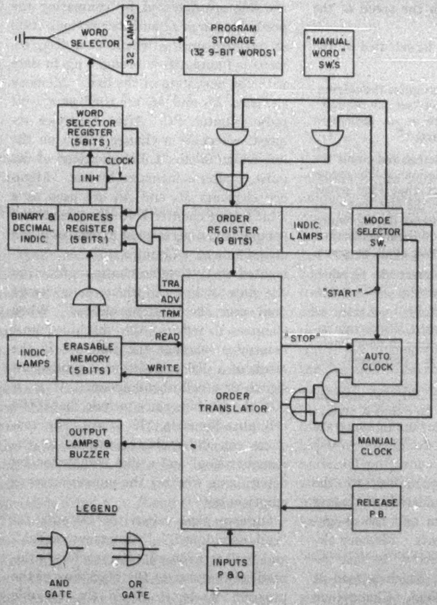


Fig. 11. General block diagram

Of particular importance is the address register. When either the *TRA* or the *TRM* lead is activated by the order translator, the address register is used to store information coming from the order register or from the erasable memory, respectively. When the *ADV* lead is activated, the address register acts as a counter and increases its contents by one.

Observe that there exists a loop going from the address register to the word selector register, the word selector, the program storage, the order register, the order translator, and back to the address register. If information were allowed to flow through this loop continuously, improper operation would result for the following reason. Upon arrival of a clock pulse, the order translator would cause a new address to appear in the address register. This address would be immediately transmitted to the word selector causing a new word to be read out of the program storage and transmitted to the order register. Since the clock pulse that generated the new address would still be present, the new word would be executed immediately. This, in turn, would generate a second new address, and so on, until the clock pulse disappeared. To insure an orderly processing of the orders, the loop previously mentioned is opened during the presence of the clock pulse by the inhibiting gate between the address register and the word selector register. Thus, the new address can reach the program storage only after the removal of the clock pulse. The new order is read into the order register, but it cannot be executed until the arrival of the next clock pulse.

Some Examples of Programs

The versatility of SPUD is sufficient to justify the term "universal." Not only is a fairly large variety of orders made available to the programer but, in addition, the programer can make the most out of the limited program storage capacity by utilizing multiple-action orders. As shown later in this section, universal program formats can be developed for certain types of combinational and sequential problems.

Twenty-five different types of orders are listed in Table I. If all the possible ways of assigning binary values to the *X*'s are considered, a total of 497 different

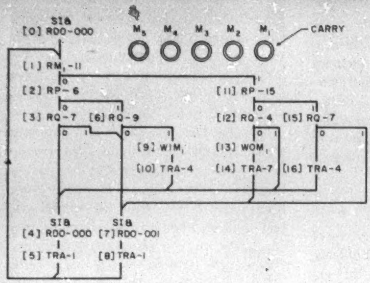

SI8
[0] RDO-000
M_5 M_4 M_3 M_2 M_1
CARRY
[1] RM₁-11
0
1
[2] RP-6
[11] RP-15
0
1
0
1
[3] RQ-7
[6] RQ-9
[12] RQ-4
[15] RQ-7
0
1
0
1
0
1
0
1
[9] WIM₁
[10] TRA-4
[13] WOM₁
[14] TRA-7
[16] TRA-4
SI8
[4] RDO-000
[5] TRA-1
SI8
[7] RDO-001
[8] TRA-1

Fig. 12 (above). Program for serial addition

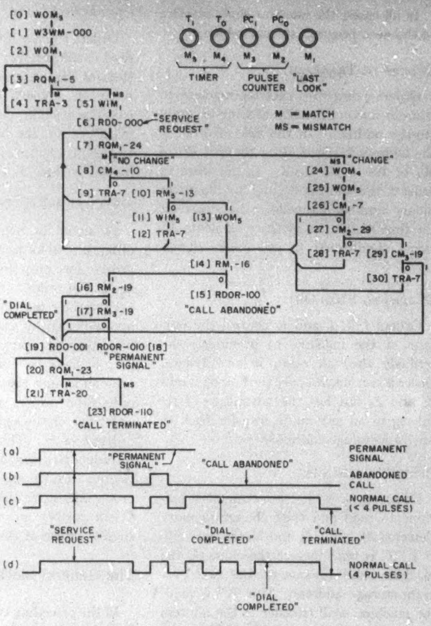

[0] WOM₅
[1] W3WM-000
[2] WOM₁
[3] RQM₁-5
M
MS
[4] TRA-3
[5] WIM₁
[6] RDO-000
"SERVICE REQUEST"
T_1 T_0 PC_1 PC_0
M_5 M_4 M_3 M_2 M_1
TIMER
PULSE COUNTER
"LAST LOOK"
M = MATCH
MS = MISMATCH
[7] RQM₁-24
M
"NO CHANGE"
MS
"CHANGE"
[8] CM₄-10
0
1
[9] TRA-7
[10] RM₅-13
0
1
[11] WIM₅
[12] TRA-7
[13] WOM₅
[24] WOM₄
[25] WOM₅
[26] CM₁-7
1
0
[27] CM₂-29
0
1
[28] TRA-7
[29] CM₃-19
0
1
[30] TRA-7
[14] RM₁-16
1
0
[15] RDOR-100
"CALL ABANDONED"
[16] RM₂-19
1
0
[17] RM₃-19
1
0
"DIAL COMPLETED"
[19] RDO-001
RDOR-010 [18]
"PERMANENT SIGNAL"
[20] RQM₁-23
M
MS
[21] TRA-20
[23] RDOR-110
"CALL TERMINATED"
(a)
(b)
(c)
(d)
"PERMANENT SIGNAL"
"CALL ABANDONED"
PERMANENT SIGNAL
ABANDONED CALL
NORMAL CALL (< 4 PULSES)
"SERVICE REQUEST"
"DIAL COMPLETED"
"CALL TERMINATED"
NORMAL CALL (4 PULSES)
"DIAL COMPLETED"

Fig. 13 (right). Program for a line register

words is obtained. Thus, there are $497^{32} \cong 10^{86}$ different ways of filling out a program card; of these, only an extremely small percentage will yield meaningful programs. Even so, there should be at least several hundred thousand meaningful programs.

The capacity of the program storage is sufficiently small to be easily grasped in its entirety; at the same time, it is large enough to make possible many fairly complicated programs. The remainder of this section will be devoted to the discussion of some of the programs that have been written for the machine.

SERIAL ADDITION (FIG. 12)

This program will add two binary numbers of any length. Upon request from the machine, corresponding bits of the two numbers are set by the operator on the input switches P and Q, starting with the least significant bits. The carry is stored in memory element M_1; the sum is given out in serial form on the Z lamps by means of a 000 output for "0" and a 001 output for "1." At the beginning of the program, the machine requests the first pair of bits and, at the same time, gives out an output 000. This output is meaningless at this time, but it is impossible to light the *SI* lamp for an input request without lighting also the *RDO* lamp. The machine then reads the carry and the bits of the addend and augend, changes the carry, if necessary, gives the appropriate output, and requests a new input. This routine is repeated over and over again, until all the bits have been fed in.

LINE REGISTER

In this program, the "0" and "1" positions of input switch Q simulate the "on-hook" and "off-hook" conditions of a hypothetical telephone line. Up to four pulses may be "dialed" on this line by operating switch Q up and down at a rate that is compatible with the speed of the machine.

The requirements to be satisfied by the program are

1. The machine shall recognize the origination of a "call" and give a "service request" output, which will serve as an indication that "dialing" can be started.
2. The machine shall detect and count the dial pulses. Dial completion shall be recognized either by the fact that four pulses have been counted, or by the fact that some interval of time has elapsed during which the line has been in the "1" or off-hook state. A "dial completed" output shall be given.
3. The machine shall detect the termination of a "call" and give a "call terminated" output.
4. Once a call is originated, if a certain interval of time elapses before dialing starts, the machine shall give a "permanent signal" output.
5. If a "call" is abandoned, the machine shall detect this by the fact that a certain interval of time has elapsed during which the machine has been in the "0" or on-hook state.

The lower part of Fig. 13 illustrates the various cases to be considered; the upper part gives the program and the assignment of memory elements. Memory element M_1 is used to record the last observed state of line Q, which is read at discrete intervals of time. Changes on the line are detected by comparing the present state of Q and its previous state as recorded in M_1; whenever a disagreement is found, M_1 is brought up to date with the new state of the line. Memory elements M_2 and M_3 are used as a 2-bit pulse counter *PC*. After a service request, whenever a change occurs on the line from "0" to "1," the content of the pulse counter is increased by one. Memory elements M_4 and M_5 are used as a 2-bit timing counter T to recognize a dial completion, a permanent signal, or a call abandonment. Counter T is incremented every time no change is found in the state of line Q; whenever a change does occur, the counter is cleared. When counter T recyles, the machine must determine whether the time-out is the result of a dial completion, a permanent signal, or a call abandonment. If M_1 is "0," the machine can conclude that it is a call abandonment. If $M_1 = 1$, the machine can differentiate between a permanent signal and a dial completion by determining whether the pulse counter is empty or not

Observe that, after the outputs for "call abandoned," "permanent signal," and "call terminated," the machine automatically returns to the beginning of the program where it clears the erasable

Vol. 7
Pt. 1
1959

x y z w	f(x,y,z,w)	Corresponding word location
0 0 0 0	f_0	16
0 0 0 1	f_1	17
0 0 1 0	f_2	18
0 0 1 1	f_3	19
0 1 0 0	f_4	20
0 1 0 1	f_5	21
0 1 1 0	f_6	22
0 1 1 1	f_7	23
1 0 0 0	f_8	24
1 0 0 1	f_9	25
1 0 1 0	f_{10}	26
1 0 1 1	f_{11}	27
1 1 0 0	f_{12}	28
1 1 0 1	f_{13}	29
1 1 1 0	f_{14}	30
1 1 1 1	f_{15}	31

(A)

M_5	M_4	M_3	M_2	M_1
1	x	y	z	w

[0]	SI & RDO-100	[16]	RDOR - 0$f_0$0
[1]	WPM$_4$	[17]	RDOR - 0$f_1$0
[2]	WQM$_3$	[18]	RDOR - 0$f_2$0
[3]	SI & RDO-001	[19]	RDOR - 0$f_3$0
[4]	WPM$_2$	[20]	RDOR - 0$f_4$0
[5]	WQM$_1$	[21]	RDOR - 0$f_5$0
[6]	WIM$_5$	[22]	RDOR - 0$f_6$0
[7]	TRM	[23]	RDOR - 0$f_7$0
[8]		[24]	RDOR - 0$f_8$0
[9]		[25]	RDOR - 0$f_9$0
[10]		[26]	RDOR - 0f_{10}0
[11]		[27]	RDOR - 0f_{11}0
[12]		[28]	RDOR - 0f_{12}0
[13]		[29]	RDOR - 0f_{13}0
[14]		[30]	RDOR - 0f_{14}0
[15]		[31]	RDOR - 0f_{15}0

Meaning of Outputs

100 = "Set x in P and y in Q"
001 = "Set z in P and w in Q"
000 = "f(x,y,z,w) = 0"
010 = "f(x,y,z,w) = 1"

(B)

Present State of Memory	Input P 0	Input P 1
0 0 0	000,--	m_{16}, n_{17}
0 0 1	m_2, n_3	m_{18}, n_{19}
0 1 0	m_4, n_5	m_{20}, n_{21}
0 1 1	m_6, n_7	m_{22}, n_{23}
1 0 0	m_8, n_9	m_{24}, n_{25}
1 0 1	m_{10}, n_{11}	m_{26}, n_{27}
1 1 0	m_{12}, n_{13}	m_{28}, n_{29}
1 1 1	m_{14}, n_{15}	m_{30}, n_{31}

(A)

$$M_5 \quad \underbrace{M_4 \quad M_3 \quad M_2}_{\text{Memory State}} \quad M_1$$

P O

[0]	WPM$_5$	[16]	W3WM - m_{16}
[1]	TRM	[17]	RDOR - n_{17}0
[2]	W3WM - m_2	[18]	W3WM - m_{18}
[3]	RDOR - $n_3$0	[19]	RDOR - n_{19}0
[4]	W3WM - m_4	[20]	W3WM - m_{20}
[5]	RDOR - $n_5$0	[21]	RDOR - n_{21}0
[6]	W3WM - m_6	[22]	W3WM - m_{22}
[7]	RDOR - $n_7$0	[23]	RDOR - n_{23}0
[8]	W3WM - m_8	[24]	W3WM - m_{24}
[9]	RDOR - $n_9$0	[25]	RDOR - n_{25}0
[10]	W3WM - m_{10}	[26]	W3WM - m_{26}
[11]	RDOR - n_{11}0	[27]	RDOR - n_{27}0
[12]	W3WM - m_{12}	[28]	W3WM - m_{28}
[13]	RDOR - n_{13}0	[29]	RDOR - n_{29}0
[14]	W3WM - m_{14}	[30]	W3WM - m_{30}
[15]	RDOR - n_{15}0	[31]	RDOR - n_{31}0

(B)

Fig. 14 (left). Universal program for a 4-variable combinational problem

Fig. 15 (above). Universal program for a sequential problem with one input and eight memory states

memory before handling a new "call" on line Q.

4-VARIABLE COMBINATIONAL PROBLEMS

It is possible to write a general type of program based on the structure of the truth table shown in Fig. 14(A) for a Boolean function of four variables *x*, *y*, *z*, and *w*. Each of the f_i entries may be either "0" or "1." Fig. 14(B) shows the program and the assignment of the memory elements. At address 0, the machine requests the operator to set the binary variables *x* and *y* in input switches *P* and *Q*, respectively. The machine copies the values of *P* and *Q* into memory elements M_4 and M_3 (addresses 1 and 2), and asks the operator to set the remaining variables *z* and *w* in *P* and *Q* (address 3). The values of *P* and *Q* are copied then into memory elements M_2 and M_1 (addresses 4 and 5). The erasable memory contains now the complete 4-variable combination for which an output is required. When a "1" is prefixed to this combination (address 6), an address is formed which uniquely identifies one of the word locations 16 to 31. Thus, there is a one-to-one relationship between these locations and the 16 entries of the truth table. At address 7, the machine is instructed to transfer to the address stored in the erasable memory. Each of the word locations 16 to 31 contains an *RDOR* order which will give as an output the value of the function to be realized for the corresponding 4-variable combination. The machine automatically transfers to address 0 and is ready to handle a new combination of variables. Observe that the universal program of Fig. 14(B) can be adapted to any specific 4-variable Boolean function merely by assigning binary values to the f_i's.

SEQUENTIAL PROBLEMS

Fig. 15(A) shows the state table for a sequential problem with one input and eight memory states. At the intersection of a row and a column, there is a double entry (m_i, n_{i+1}), where m_i stands for a 3-bit combination defining the "next state of the memory",[2] and n_{i+1} stands for a 2-bit output combination. The only restrictions are

1. When the "present state of the memory" is 000 and the input is 0, the next state of the memory" is still 000 and no output is required.
2. Only four output combinations are available.

Within these restrictions, which result from the limited capacity of the machine, it is possible to write a general type of program based only on the structure of the state table. Such a program is shown in Fig. 15(B) together with the assignment of the memory elements. The first two steps of the program instruct the machine to copy the input *P* into M_5 and to transfer to whatever address happens to be in the erasable memory. Observe that when the power switch is first turned on, all the memory elements

M_1 to M_5 are in the "0" state. Therefore, as long as the input P is "0," the machine will shuttle between addresses 0 and 1. When the input P is changed to "1," the machine will transfer to address 16 (10000 in binary form), where it will write the "next state of the memory" m_{16} in M_4, M_3, and M_2. Assume, as an example, that $m_{16}=011$. In this case, the machine will write 011 in M_4, M_3, and M_2, and will advance. At address 17, the machine sounds the buzzer, reads out the 2-bit output n_{17}, and locks itself up. Even though the *SI* lamp does not go on, the operator must interpret the buzzer as a request to set the next input in P. When the operator presses the release button, the machine automatically transfers to address 0, where it copies P into M_5. It advances then to address 1 where it is instructed to transfer to the address stored in the erasable memory. If $P=0$, the machine transfers to 6 (00110 in binary form); if $P=1$, the machine transfers to 22 (10110 in binary form). The "next state of the memory" will be written in M_4, M_3, and M_2, an output will be given, and the routine will start all over again.

A program card punched in accordance with the universal program of Fig. 15 can be adapted to any specific problem merely by entering the required 3-bit m_i combinations and the 2-bit n_{i+1} combinations. The same program, of course, can be applied to a sequential problem requiring fewer states. In a similar fashion, a general type of program can be written for a sequential problem with two inputs and four memory states.

Conclusions

The machine described can facilitate considerably the teaching of programing principles for information-processing machines. It offers the student an opportunity to develop many interesting programs and to test their inplementation in a step-by-step fashion. The machine provides also an abstract model that can be a useful tool by itself.

References

1. The Design of Switching Circuits (book), **W. Keister, A. E. Ritchie, S. H. Washburn.** D. Van Nostrand Company, Inc., Princeton, N. J., 1951.

2. A Method for Synthesizing Sequential Circuits, **G. H. Mealy.** *The Bell System Technical Journal*, New York, N. Y., vol. 34, Sept. 1955, pp. 1045–79.

Experience with Long-Distance Television Fields Used for Retransmission

W. L. BRAUN
MEMBER AIEE

IT HAS BEEN KNOWN for some time that long-distance (100–150 miles) very-high frequency propagation exists, and is capable of wide-band performance. This paper follows the empirical investigation of these fields to their ultimate use as signal sources for a community antenna system (CATV) and a television station (WSVA-TV).

The engineering staff of WSVA, Harrisonburg, Va., investigated frequency-modulated signal reception from April 1946 to December 1947. The results of these tests were to be applied to the siting problem of WSVA-FM. These field measurements indicated the existence of signals well beyond the limits generally assumed at the time. Various projects of high priority prevented the continuation of this project at that time. This program was reopened in August 1949 with the addition of television signal observation. Earlier conclusions as to the presence of these long-distance fields were substantiated. It was found that the stability of these signals was far greater than had been expected. No serious attempt was made to make a scientific investigation of the nature of these fields, except to investigate them by observation.

Harrisonburg is located in a natural bowl, behind two mountain ranges of the Appalachian chain, approximately 3,500 feet below horizon. The nearest sources of television signals in 1950 were Richmond, Va., and Washington, D. C. A few of the local residents attempted to receive these stations directly, but the results were unsatisfactory. It was felt that this situation justified the installation of a community antenna system (CATV). Field research was commenced immediately, following the guide of the information that had been accumulated previously. Since the nearest mountain range was approximately 8 airline miles distant, it seemed advisable to search the local area about the city to see if satisfactory reception could be found. It was discovered immediately that any site had to be explored in detail to evaluate it, as the fine-grain variation of the signal on a given site was quite substantial. It had been presumed that the greatest elevation of site would produce the largest signal. This did not prove to be an infallible rule. It was proved equally incorrect to assume that the signal increased with increasing height to the free space value. The actual siting situation can be summarized as follows:

1. The foreground clearance to the nearest obstruction must be reasonably uniform.
2. The shape of the foreground terrain slope is important.
3. The signal increases with height in Fresnel-zone fashion to a maximum, then minimum, then to a new maximum, but with each new maximum decreasing in intensity over the first.
4. The signal maximum is localized for a given site in height, width, and depth, and has to be located empirically.
5. The signal maximum outlined in items 3 and 4 is frequency-sensitive: 12- to 9-foot average for channels 7 through 13 and 40- to 35-foot average for channels 2 through 6.
6. Due to item 4, the position and size of the antenna aperature to be used must not exceed the signal "capsule."
7. The time-fading factor within this signal capsule is better than in an area outside it.
8. This is a small area phenomena, and since the time-fading factor is superimposed over the top of the exploration, great care must be taken to repeat the measurements enough times to assure the siting of the optimum location.
9. The signal gain possible by this technique is substantial, often as much as 15 db (decibels).

The optimum location could be gen-

Paper 59-674, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Baltimore, Md., May 19–21, 1959. Manuscript submitted September 19, 1958; made available for printing March 26, 1959.

W. L. Braun is with WSVA Shenandoah Valley Broadcasting, Inc., Harrisonburg, Va.

The author wishes to acknowledge the assistance of the following engineers who participated in the program described in this paper: Thomas B. Jones, J. Maxwell Foley, Ira T. Lowe, Robert C. Jackson.

erally described as the forefront of an elevated location, just before the crest of the elevation, with the slope of the forefront of this location approximating any of the curves possessing a focus accessible by the antennas. The further this condition prevailed before the major obstruction, the more nearly optimum location for one channel on a given bearing did not necessarily fit for another channel on the same bearing. In one case it was nearly 300 yards away.

Table I. Circuit Transmission Losses

Channel No.	Station Location	Length of Path, Air Miles	Path Loss, Db[1]
4	Washington, D. C.	90.0	135
5	Washington, D. C.	85.0	128
7	Washington, D. C.	90.0	145
8	Petersburg, Va.	98.5	149
9	Washington, D. C.	89.5	141
10	Roanoke, Va.	142.0	148
12	Richmond, Va.	93.5	135

The signal intensity at this optimum site was too low to deliver a satisfactory picture with a dipole antenna. It was felt that at least 20 db of signal improvement was necessary to achieve a satisfactory picture. A figure of 18 db was assigned to obtain as antenna gain over a dipole. The remaining gain was to be obtained by reduction of preamplifier noise figure and reduction of transmission-line loss by installing the preamplifier at the antenna terminals. All these proved to be reasonable goals.

A few interesting problems developed in the construction of antennas for this gain. This first antennas tried were 10-wavelength-per-leg horizontal rhombics. The first of these antennas performed satisfactorily, but did not deliver the calculated gain by 25% (voltage). This was found to be due to the excessive size of the rhombic exceeding the signal capsule. A longer rhombic was tried with 20 wavelengths per leg, but the major lobe split, and the resultant gain was very low. An assessment of the problem indicated the need for a compact antenna with gain in all three co-ordinate directions. A preliminary paper design indicated that arrays of yagi antennas, close spaced, would solve the problem. The first antenna was an array of 8–8-element yagi antennas with 0.62-wavelength colinear spacing and 0.75-wavelength vertical spacing (entire array was two high and four wide). This antenna produced its calculated gain when propagation conditions were good, but deteriorated somewhat when propagation conditions were poor. This was determined by making interarray phase measurements on the arriving signal. By very careful location of the array in the signal capsule, satisfactory results were obtained most of the time. A voltage standing wave ratio measurement of the completed array indicated a maximum value of 2 to 1 without any compensation other than simple matching. By the use of very simple networks, the voltage standing wave ratio was reduced to 1.5 to 1.

Two years of experience, until June 1953, indicated the following:

1. General level of night signal was 6–10 db higher than at the day signal.
2. The signal was least consistent about sunset.
3. Heavy overcast accompanied by high relative humidity always increased the median signal and reduced the variations about the median.
4. Occasionally the propagation would go through quick and deep fades for very short periods of time (not to be confused with aircraft flutter).
5. Reliability of circuits was excellent. Down-time was less than 0.2% on the poorest circuit.
6. Rhombics produced a characteristic "transient distortion," making their use less than desirable for this purpose.

The last item proved to be a difficult one to accept, but several years later this was found to be entirely true, and due in part to the phase shift of component parts of the signal over the length of the rhombic.

When WSVA-TV was constructed, it seemed advisable to use this system of signal source, as the television station was to be located on a very remote mountain top, 60 miles and two mountain ranges away from the nearest American Telephone and Telegraph interconnection point. Initial experiments on the television mountain top indicated a similar signal situation to that experienced at the community antenna site just under discussion, except that the median field was about 15 db higher. This site was 2,000 feet higher and 11 miles closer to the stations to be received than the CATV receiving site. It is interesting to note that the same signal capsule existed on the mountain site, but was less pronounced in the effect on the location of such a capsule. Since the antennas had to be located where the best signal existed, it was necessary to preamplify the signal and relay it to the transmitter building through coaxial cable type *RG11/U*. As soon as the system was activated with the transmitter operating, a series of problems developed as the result of spurious outputs from the transmitter, channel 3, and the spurious responses of the receivers operating on channels 4–10 and 12. By straightforward interference elimination (filters, etc.) these problems were eliminated. It was most difficult to remove the intermodulation components radiated from the transmitter, but it was done. Due to the greater median field intensity at the transmitter site, it was possible to reduce the antenna arrays to a simple stack 10-element yagi array for each channel.

After 4 years' experience with this arrangement, several factors appear as significant:

1. Mode of propagation is diffraction over the mountain obstructing the path. This is true even though the mountain ranges are heavily wooded.
2. Signal quality is excellent, with very little variation about the median value.
3. Signal quality necessary for retransmission must be much higher than that for CATV work, as the equipment differentiates the leading edges of pulses for clamping pulses. Poor transient response will render such equipment unusable.
4. Spurious responses of receivers are extremely difficult to cope with. It is better to use crystal-controlled converters and operate the receivers on a channel free of such spurious components.
5. Circuit reliability is as good as 99.95% on a yearly cumulative basis. Only difficulty occurs momentarily with "sporodic *E*."
6. Antenna siting should be done with the same care as with a less elevated location. The same gains can be realized.

No analysis has been made in depth to determine the type of distortions present for use with color (National Television Systems Committee standard). A preliminary observation indicates that, superficially, these circuits are usable for color. Investigation has been under way for some time to specify the exact nature of the distortions, if any, that exist in the transmission path.

Conclusions

The data and experience presented in this paper have shown the practicality of using long-distance "diffraction" fields for CATV and television broadcast use. The significant factors surrounding the use of such fields are as follows:

1. Local investigation of terrain effects is necessary to obtain optimum antenna siting.

2. Antenna gain should take place in the smallest possible physical volume practical, and with minimum height of array.

3. Reliability of such signals is excellent.

4. More research needs to be done to determine exact nature of transmission-path distortion for color. It may be acceptable as it is.

For those in the propagation field, a listing of circuit transmission losses is given in Table I.

References

1. TRANSMISSION LOSS IN RADIO PROPAGATION, Kenneth A. Norton. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, no. 1, Jan. 1953, pp. 146–52.

2. PREDICTION OF THE NOCTURNAL DUCT AND ITS EFFECT ON UHF, L. J. Anderson, E. E. Gossard. *Ibid.*, pp. 136–39.

3. LARGE REDUCTION OF UHF TRANSMISSION LOSS AND FADING BY THE PRESENCE OF MOUNTAIN OBSTACLE IN BEYOND-LINE-OF-SIGHT PATHS, F. H. Dickson, J. J. Egli, J. W. Herbstreit, G. S. Wickizer. *Ibid.*, no. 8, Aug. 1953, pp. 967–69.

The Use of Automatic Programing Techniques for Solving Engineering Problems

J. T. CARLETON
MEMBER AIEE

N. CHACKAN
ASSOCIATE MEMBER AIEE

T. W. MARTIN
NONMEMBER AIEE

THERE are four large-scale digital computers at Westinghouse Electric Corporation employed in making engineering calculations: one International Business Machine Corporation (IBM) *705* and three IBM *704*'s (Table I). This paper summarizes the experience with automatic coding systems at these four installations.

Analytical Department

The Analytical Department located at the East Pittsburgh, Pa., Works has an IBM *704* digital computer with a 8,192-word main core storage, 8,192 words of drum storage, and eight magnetic tape units. A peripheral tape-to-printer unit, card-handling equipment, and key punches are also available.

Most of the programs written for this machine are used to solve engineering problems for the Large Rotating Apparatus Department, the Industry Engineering Department, Atomic Power Department, Kansas City, Kans., Works (aviation gas turbines), Lima, Ohio, Works (small motors), South Philadelphia, Pa., Works (steam turbines), and for various Westinghouse customers such as electrical utility companies. A small amount of data processing and engineering problems from other sources is also handled.

Most of the Analytical Department programs, once written, have a rather short life, the charges for production on a job frequently being less than the program development cost. An obvious conclusion is that the ease of writing and debugging a program are far more important to an installation with this load characteristic than the efficiency of the final program.

There is a staff equivalent to 50 full-time programers generating programs at this installation. The work of this staff results in an average of 150 hours of billed computer time per month with a breakdown of 100 hours to prepare and test programs and 50 hours production on completed jobs. The majority of those who make up this staff of 50 are professional programers, although about 16% of the work is produced by casual programers working on an open-shop basis using the FORTRAN automatic coding system. A breakdown of the remaining 84% shows that 10% is coded using FORTRAN and 84% using the symbolic assembly program (SAP).

A sample of six programs written using the FORTRAN system was analyzed to determine the cost of preparation with the results compared to cost estimates for the same programs using the SAP system. The estimates of man-hours required using SAP were made on a basis of factors which have been extensively used in estimating costs in the department. Once the number of man-hours was reached, the estimate of machine time required for program assembly and debugging was made by using a ratio of 2 machine-hours per man-month. Of the six programs in the sample, three were calculations easily coded by FORTRAN for an 8,192-word *704*. These three were written by a full-time programer with previous experience using FORTRAN. Another was written jointly by a full-time programer and a design engineer with little FORTRAN experience. The last two of the sample were written by men with little previous experience in FORTRAN. One of these samples presented the further complication that the desired result could not be completely accomplished with FORTRAN and required modifications in the finished machine language program.

The following lists the results of the analysis. None of the production on the jobs is included in these figures; all time listed was spent in generating and testing the program.

	Man-Hours	Machine-Hours
FORTRAN (actual)	883	15.8
SAP (estimate)	2,541	28.9

Since these programs were produced, a new version of the compiler has been made available. This system, a modified version of FORTRAN II, has resulted in a minimum savings of 30% in compiling time, with cases on record of 50% savings. Figures available from accounting records of the department indicate that of the 15.8 machine-hours, 7.6 were consumed by the compiling process. Applying the 30% savings to this amount of machine time, a new table was constructed as follows:

	Man-Hours	Machine-Hours
FORTRAN (modified)	883	13.5
SAP (estimate)	2,541	28.9

It can be seen that an estimated savings of 63.2% in man-hours (elapsed time if it is assumed that one man did all the work) and 53.3% in machine-hours could be realized by using the FORTRAN system. From available published figures, representative costing rates per *704* hour

---

Paper 59-641, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Baltimore, Md., May 19–21, 1959. Manuscript submitted February 19, 1959; made available for printing April 29, 1959.

J. T. CARLETON and T. W. MARTIN are with the Westinghouse Electric Corporation, East Pittsburgh, Pa.; N. CHACKAN is with the Westinghouse Electric Corporation, Sharon, Pa.

Table I. Large-Scale Digital Computers Used for Engineering Calculations

Location	Computer	Type of Work	Engineering Usage, Shifts	Production Time, Hours per Month per Programer	Program Testing Time, Hours per Month per Programer	Cost Ratio, Preparation to Production
Analytical Department, East Pittsburgh, Pa.	IBM 704	Service Bureau	1+	1	2	5
Air Arm and Electronics Divisions, Baltimore, Md.	IBM 704	Military systems performance and design calculation	1+	5	3	2
Atomic Power Division, Pittsburgh, Pa.	IBM 704	Atomic plant performance calculation	4+	30	$3^1/_2$	1/5
Transformer Division, Sharon, Pa.	IBM 705	Transformer design data processing	0.4	3	6	4

and per programer hour were determined. Using these rates, the cost of generating the programs using FORTRAN would be $13,555 compared with an estimated cost using SAP of $35,525. This represents an estimated savings of 61.9% in the preparation cost of the jobs.

These figures were applied to the load of the Analytical Department computer with the following assumptions:

1. Fifty per cent of the work of the department is either unsuitable for FORTRAN coding or is already being accomplished in FORTRAN. This assumption implies that 24% of the work is unsuitable for FORTRAN since 26% is being written using that system as pointed out previously.
2. The remaining 50% is such that the 2-to-1 ratio between debugging hours and production hours still holds.
3. A program resulting from FORTRAN is 20% less efficient than one generated by a skilled programmer using SAP. This appears from experience to be a more-than-generous allowance for the inefficiencies of a FORTRAN program. It also could be pointed out that careful coding is necessary to produce an optimum SAP program and would thus increase the number of man-hours required to produce the final program.

Table II represents the effect of applying the figures from the foregoing tables.

It will be noted that this results in a decrease of 21.65 machine-hours billed to customers per month; so in evaluating the cost savings, a machine overtime rental rate per *704* hour will be used. Using this figure, the same amount of work could be accomplished with a staff of 15 fewer programers and at a savings of 21% present total expenses.

FORTRAN SERVICE

To broaden the base of computer usage and to "computer-orient" the man with the problem, a FORTRAN Service has been inaugurated. The compiler was simplified by prohibiting the use of the more sophisticated statements which experience indicated were common sources of programing errors. Reference material in the form of a FORTRAN service manual (summarized in Fig. 1) and a 2-hour course of instruction were arranged. These, together with occasional assistance from a FORTRAN Service Group, have been all that is necessary to make the power of the IBM *704* available to the engineer with the problem. Preparation of the cards, operation of the computer, and other details are handled by the FORTRAN Service Group so that the user need have no knowledge of computer operating procedure.

More than 200 engineers in the East Pittsburgh area attended the 2 hours of instruction. Subsequently, a survey was made which yielded the following results:

1. Twenty-two per cent of those attending wrote programs.
2. Of those who wrote programs, one-third did not attempt to compile. Of the remaining two-thirds, all produced useful programs.
3. All those interviewed felt capable of programing a simple problem.

Perhaps the most significant development resulting from the introduction of the FORTRAN Service has been the development of a number of extremely competent programers from among our engineers. After mastering simplified FORTRAN, many have proceeded on their own initiative to use the more sophisticated statements to the extent that they now have at their disposal the full capability of the *704* and FORTRAN.

Air Arm, Electronics Divisions

The IBM *704* at Air Arm Division, Baltimore, Md., jointly used by Air Arm and Electronics Divisions, has a main core storage of 8,192 words, drum storage of 8,192 words, eight tape units, and peripheral equipment. Its principal use is in the determination of military systems performance, detailed design of electronic components, and reduction of flight test data. Programs once written have a moderate life, with the production running time slightly greater than the program writing and testing time.

Table II. Application of Modified Fortran Figures

	Machine, Production Hours	Machine, Debugging Hours	Programer Hours, 176 Hours Per Man-Month
Present time	50	100	8,800
Estimated change	+ 5	−26.65	−2,781
Result	55	73.35	6,019

This computer was placed in operation in October 1957, coincident with the availability of FORTRAN. Although the personnel had had considerable experience with SAP on the Analytical department machine, FORTRAN was almost immediately adopted. By December 1957 the bulk of the programing was done in FORTRAN, and at present approximately two-thirds of the programing effort is in FORTRAN.

Several changes in FORTRAN were made, including:

1. A binary tape loader.
2. Print-out of octal instruction locations to facilitate making corrections and modifications.
3. Expansion of library routines to include Bessel functions, solution of simultaneous equations, and solution of nonlinear simultaneous differential equations.

This installation has not made a detailed comparison of conventional programing methods with automatic programing techniques; however, these facts are significant.

Prior to the installation of this computer, a study was made of IBM *704* facilities doing similar work. This study indicated that 40 programers would be required for one shift operation. Using FORTRAN, only 25 programers have been required.

The solution of a 25th order nonlinear differential equation was required to study an air-to-air missile. Five programers required 6 months to get a working program using SAP. Using FORTRAN

SYMBOL	OPERATION	EXAMPLES OF WRITING NUMBERS CORRECTLY	MANUSCRIPT CONVENTIONS LETTERS	MANUSCRIPT CONVENTIONS NUMBERS
**	Exponentiation (Raise to Power)	17.	I	1
*	Multiply	3.1416	O	0
/	Division	.967 * 10. ** (-23.)	U	
+	Addition	17. * 6.3 ** .823	Z	
-	Subtraction		V	

FORTRAN	FUNCTIONS
SINF (X)	Sine X
CØSF (X)	Cosine X
SQRTF (X)	Square Root of X
NLØGF (X)	Log_e X
ARTNF (X/Y)	Tan^{-1} (X/Y)
ARTNF (X)	Tan^{-1} X
EXPF (X)	e^X
TANHF (X)	Hyperbolic Tan X
TBLF (A,B,A1)	Table Lookup, B(A1)
ABSF (X)	X

INPUT INFORMATION

STATEMENTS	
Read 30000	Alphabetic Input for Immediate Printing
Read 31001,A,B,...	Data Input, 5 Items/Card

OUTPUT INFORMATION

Print 30000	Print Alphabetic Input Immediately after Read-In
Print 32000	Obtain New Sheet of Paper
Print 32005,X,Y,...	Print up to Five Columns of Numbers
Print N	Print Alphabetic Line from Format Statement N at Bottom of Coding Form

EXAMPLES

FORTRAN STATEMENT	MEANING
A + B + C - X	A + B + C - X
A * B * C	ABC
(A/B) * SINF (C)	(A/B) Sine C
A ** B * C - NLOGF (D)	$(A^B C) - \text{Log}_e D$
A + B/C + D ** E * F - G	$A + (B/C) + (D^E F) - G$
A - ((B/C) * (D **E)) + G	$A - ((B/C)(D^E)) + G$
A - B/C * D ** E + G	$A - ((B/C)(D^E)) + G$
A ** B ** C	Not Allowed

When in Doubt Use Parentheses ( ).

GENERAL INFORMATION

1. The name of a variable must not begin with I,J,K,L,M,N or a number.
2. Statement numbers are written in any column from 1 through 5. A statement number can be any number between 1 and 9999. They do not have to be in order.
3. When using C for comments, C is always written in Column 1. A line with C in column 1 will be ignored by Fortran.
4. If a statement is too long to fit on a single line it can be continued over as many as 9 additional lines. Continuation lines must be numbered consecutively from 1 to 9 in Column 6.
5. Arithmetic operations are performed in the following order: Exponentiation, Multiplication and Division, Addition and Subtraction. Operations are performed from left to right.

Fig. 1. Simplified FORTRAN summary

and the Integration with Controlled Error Library Routine, two programers were able to get a similar program working in 6 weeks. Although learning should be credited with part of this reduction, the use of FORTRAN was certainly the dominant factor.

Atomic Power Division

The Atomic Power Division has an IBM *704* with main core storage of 32,768 words, magnetic drum, tapes, and peripheral equipment.

Most of the programs run on this machine involve the solution of partial differential equations associated with calculating atomic power plant performance. The programs are general in nature and, once written, are widely used over an extended period of time. Only 10% of the computer time is used for writing and testing programs.

Because of a large dollar expenditure in production (in excess of four shifts), the running time is of prime importance. Therefore the bulk of the programing is in SAP, very little FORTRAN being used.

Transformer Division

Transformer Division's IBM *705* is used for data processing and engineering. The main frame has a storage capacity of 40,000 digits with auxiliary drum storage of 60,000 additional digits. Ten tape units are connected to the computer in addition to a printer, card punch and card reader. Peripheral equipment consists of two tape-to-printer units. A key punch is located in the main computer room for instant accessibility to computer users.

Transformer Division's major engineering problem is the design of transformers. This is divided into three basic groups; design of shell-form transformers 12,000 kva and up, design of core-form transformers, and design of reactors. Each design problem is treated separately and independently of the others. The computer solutions to these and other engineering problems require the full-time effort of five engineers plus approximately five nonengineers. Since the main memory is insufficient for long detailed programs such as transformer design, these programs are subdivided into many parts each of which is located on tape or drum and called in when needed. Thus, a program is capable of an unlimited number of instructions. Results of calculations are stored in memory, on tape, and on drum. Electrical and design specification data and forms are prepared on tapes for peripheral printout.

The computer operates two 8-hour shifts. Of this total, approximately 20% is devoted to the Engineering Department. At present the total engineering time is approximately 65% debugging, 35% production.

Conclusions

Automatic programing methods such as FORTRAN provide a powerful tool in the hands of experienced programers and will play a major economic role at installations where the cost of writing and checking programs is equal to or greater than the production cost.

A simplified version of FORTRAN supported by a service organization to handle the mechanics of operation and scheduling is an excellent method for making the facilities of the *704* available to a large number of engineers. It also provides an inexpensive and effective method of training engineers to become competent programers.

The ease of programing provided by FORTRAN allows more attention to be focused on the analysis of the problem. Improvements made by users have further increased the usefulness of FORTRAN by allowing the programer's orientation to turn even more to the problem where it belongs.

Compilers require that the computer be used to prepare the final program; however, the extra computer time is recovered with interest by a reduction in errors and a consequential decrease in computer time during the debugging phase.

The use of compilers is not restricted to large-scale computers. Small- and medium-scale computers may be effectively used for problems that are easily contained in their memory.

Appendix I. Description of Coding Systems

IBM 704

The *704* is a stored-program digital computer, i.e., an entry in the computer may be either an instruction or a number. To complicate matters, the machine uses the binary number system, a system not always easily converted from the usual decimal system. A typical instruction is CLA, clear and add, which results in clearing the accumulator and inserting the contents of a specified core storage unit. The binary code for CLA 7 is

000101000000000000000000000000000111.

An example of a number in the floating binary system is the representation of 1:

010000001100000000000000000000000000.

Other instructions or numbers are coded as different combinations of the so-called binary bits.

It is obvious that a method easier than hand coding in machine language had to be developed in order to make good use of the computer. Two such systems have been produced, SAP and FORTRAN. The SAP is a translator of symbolically coded instructions or numbers into machine language. An example might be a program to compute the circumference of a circle knowing that the radius is given at a specific location *R*. The following code when translated by SAP will produce a program which will accomplish the calculation.

```
START LDQ R       PLACE RADIUS IN MULTI-
                    PLIER-QUOTIENT REG-
                    ISTER
      FMP 2.      MULTIPLY BY 2
      STO TEMP    STORE RESULT TEMPO-
                    RARILY
      LDQ TEMP    PLACE 2R IN MULTIPLIER-
                    QUOTIENT REGISTER
      FMP PI      MULTIPLY BY π
      STO CIRC    STORE ANSWER IN CIR-
                    CUMFERENCE
      HTR         END OF CALCULATION
R                 LOCATION OF RADIUS
2.    DEC 2.
TEMP              TEMPORARY STORAGE
PI    DEC 3.1416
CIRC              LOCATION OF RESULT
      END START   END OF PROGRAM. START
                    COMPUTING WITH
                    START
```

The program causes the radius *R* to be multiplied by 2, this product to be multiplied by π, and the final result stored in CIRC. The HTR causes the computer to stop. Notice that every symbolic location must be specified in the program. A complication arises when it is desired to read in the value of the radius and print out the value of the answer. The procedure is made somewhat easier than might be expected by the use of prewritten subroutines from a library file. Other library subroutines are available such as square root, natural log, and trigonometric functions.

Another sophistication which can be included in SAP is the use of interpretive subroutines. Interpretive programing allows a programer to call for instructions which do not exist for the machine but which are translated and accomplished by the subroutines. An example of this type of programing might be the use of complex number arithmetic which the *704* cannot do directly.

The latest step towards ease of programing is the compiler, for example, FORTRAN. The compiler is a computer program which will translate a language close to English and algebra into machine language. The compiler accepts statements in the language and organizes them into a symbolic program that is fairly well optimized and which can be translated by SAP into a machine language program. To accomplish the computation of the diameter given the radius of the circle as before, one would write:

```
CIRC=2. *3.1416 *R
STOP
```

The compiler would then take over all the functions of translating into machine language and assigning all storage locations necessary. A complete program for this purpose including input and output would be

```
READ   R
CIRC  =2. *3.1416 *R
PRINT  CIRC
STOP
```

The use of the compiler can accomplish large savings in time and manpower as dramatized by the example of six FORTRAN programs. An added advantage is the ease of training programers. The engineer needs to know only the 32 types of statements of the compiler; and, in fact, he can do a great deal of programing with far fewer than the 32.

IBM 705

The *705* is similar to the *704* in that it, too, is a stored program machine. The major difference is that the *705* is a variable-length field machine using a decimal number system. The *705* is therefore more suitable to data processing where restrictions to fixed word lengths would create additional programing problems and an unjustifiable use of memory. Lengths of fields in the *705* are defined by signed fields and proper use of instructions.

A single machine instruction or program step for the *705* consists of five digits broken into two separate parts.

1. Operation part: alphabet or numeric high-order character which designates the operation to be performed.
2. Address portion: last four digits which is the address of some position in memory or some input-output device.

Occasionally a problem is programed directly in machine language thus bypassing the intermediate steps necessary to convert an automatic code to this language. To do so requires thorough knowledge of the computer plus programing care since zoning above every digit is extremely important, affecting constants as well as the instructions themselves. For relatively short problems, however, machine language may be desirable since it eliminates the assembly process.

Practically all of the Transformer Division's programs are written in symbolic language, a language that substitutes a 5-digit numeric address for an actual machine address and a 2- or 3-digit alphabetic code for the machine instruction. It is unnecessary for the programer to keep track of memory locations. He is required to specify the initial address location only if the program is to be located at some fixed place.

The *705* symbolic programing procedure does not reduce the number of instructions for a given problem. After a problem is programed in symbolic language, an assembly program converts the symbolic addresses to actual machine addresses, changes the symbolic instructions to single characters, and positions the proper zoning to identify the various accumulators and memory quadrants. As a final stage, 13 machine instructions are condensed and punched onto one 80-column card.

Another form of coding which has been used for a limited number of engineering problems is the Print I (pre-edited interpretive) routine. This is an automatic coding system of the interpretive type, the major portion of the coding and clerical work being done by the computer itself. A program consists of Print I and/or *705* symbolic instructions designed to bring in data, do arithmetic and logical operations, and produce output. After these instructions are punched on cards, the pre-edit routine of the Print I system will process the program cards and convert them to pseudo-instructions. Print I is not suitable

Table III. Basic Machine Language

Actual Location	Machine Instruction
1104	B 0004
1109	8 2013
1114	7 1169
1119	8 2021
1124	7 1179
1129	8 2039
1134	7 1189
1139	1 2714
1169	H 0000
1174	F 2609
1179	H 0000
1184	F 2619
1189	H 0000
1194	F 2629
1199	H 2609
1204	8 12$\bar{0}$4
1209	H 20$\mathring{0}$9
1214	1 0604
1219	F 2839
1224	H 2629
1229	8 12$\bar{2}$9
1234	Q 28$\mathring{3}$9
1239	1 0304
1244	F 2849
etc.	

for data processing. For short engineering problems where alphabetic data are not needed or unimportant, Print I has proved to be acceptable. A comparison between the symbolic and Print I systems will yield approximately a 30 or 40 to 1 ratio, 30 to 40 symbolic instructions to one print instruction. However, since the executive routine must always be used to interpret the pseudo-instructions, actual running time of Print I is always more than that of machine language, averaging approximately 12% more.

The present-day automatic coding system used at Transformer Division is AutoCode. This is similar to Print I except that it is a compiling system rather than an interpretive system. It is far more flexible and adapts readily to data processing.

Appendix II. Comparison of Coding Systems

As an example of the various coding systems, the following problem was coded in *705* machine language, Symbolic, Print I, FORTRAN, and in the APS III compiler. Symbolic and Print I apply to the *705*. The FORTRAN problem was run on the *704*, and the Datatron *205* was used with APS III.

Problem:

To calculate the horizontal component of flux as a function of current in a conductor and as a function of the parameters w, t, h, and k where

t = thickness of a conductor of rectangular cross section

w = width of conductor of rectangular cross section

h, k = horizontal and vertical co-ordinates of a point in the plane of the cross section

Formula:

$$\frac{wH_x}{0.2I}=1/2\left(A\ln\frac{A^2+C^2}{A^2+D^2}+B\ln\frac{B^2+D^2}{B^2+C^2}\right)+D\left(\tan^{-1}\frac{B}{D}-\tan^{-1}\frac{A}{D}\right)+C\left(\tan^{-1}\frac{A}{C}-\tan^{-1}\frac{B}{C}\right)$$

Where:

$$A=\frac{h}{t}-\frac{w}{2t},\quad B=\frac{h}{t}+\frac{w}{2t},\quad C=\frac{k}{t}-1/2,\quad D=\frac{k}{t}+1/2$$

705 Basic Machine Language and Symbolic Coding

For the *705* computer 213 instructions, 34 constants, 20 work-area regions, and a print-out area were necessary in addition to the floating point subroutines. For the floating point subroutines the maximum total number of executable instructions is 2,035, depending on what particular paths in the subroutine are taken. Programing time was 2 days; assembly time, 6 minutes; and debugging time, 1 day. This does not include time required for punching cards and the interim time between any of the various phases.

Table IV. Symbolic Language

Symbolic Location	Operation	Address	Accum.	Data, Description
01010	Set	0004	02	
01020	LOD	60000	02	Initial W/T address
01030	UNL	01090	02	
01040	LOD	60010	02	Initial K/T address
01050	UNL	01110	02	
01060	LOD	60020	02	Initial H/T address
01070	UNL	01130	02	
01080	TR	03910		Additional preparation
01090	RAD	(0000)	00	W/T, variable
01100	ST	71000	00	W/T, work area
01110	RAD	(0000)	00	K/T, variable
01120	ST	71001	00	K/T, work area
01130	RAD	(0000)	00	H/T, variable
01140	ST	71002	00	H/T, work area
01150	RAD	71000	00	W/T
01160	LOD	01160	02	
01170	RAD	60130	01	1/2
01180	TR	0604		Floating point MLPY
01190	ST	71004	00	W/2T
01200	RAD	71002	00	
01210	LOD	01200	02	
01220	RSU	71004	01	W/2T
01230	TR	0304		Floating point ADD
01240	ST	71005	00	H/T−W/2T=A

Table V. Print I Instructions

Serial	Location	Operation Code	Variable Field
01010	WOOY	REG	W001
01020	K011	REG	K001
01030	H011	REG	H001
01040	C001	FLC	+0+5
01050	T014	REG	T001
01060	ANSWR	REG	
01070		ENT	
01080		RCD	READER
01090		RPT	4, *1, 1
01100		FLO	COL01, 1, L2, W001
01110		RCD	READER
01120		RPT	6, *1, 1
01130		FLO	COL01, 1, L1, K001
01140		RPT	5, *1, 1
01150		FLO	COL07, 1, L0, K007
01160		RCD	READER
01170		RPT	6, *1, 1
01180		FLO	COL01, 1, L1, H001
01190		RPT	5, *1, 1
01220		FLO	COL07, 1, L0, H007
	etc.		

Input variables were changed automatically by the program for a large number of points. Reloading time, therefore, for additional points was nonexistent. Initial load from card reader required approximately 25 seconds. A total of 353 calculations were performed in approximately 5 minutes, each calculation having approximately 2,000 instructions plus change of variables.

Tables III and IV give the first 24 instructions of this problem in machine language and in symbolic language. The foregoing discussion applies to both methods except that no assembly time is necessary for machine language. Note that for machine language no description can be given on the cards.

Print I

Sixty-six total instructions were required in print language. The first 20 of these are shown in Table V. Running time was approximately 10% greater than machine language. Programing time was 1 day, assembly time (pre-edit), approximately 5 minutes.

Fortran

This same problem was programed for the *704* computer. The entire program is listed in the following, a total of 20 instruction statements plus eight format and dimension statements. Programing time was less than 10 minutes, excluding setting up the equations. Conversion of statements into machine language, resulted in 290 instructions, excluding subroutines. Running time for a problem of 1,008 points was 6 minutes distributed as follows:

Set up and lost motion	1–20
Compiling	2–50
Load program	0–20
Calculate 1,000 cases	1–30
Total	6–00

```
DIMENSION XHT (12), XKT (12), XWT
  (7)
WRITE OUTPUT TAPE 6, 32000
WRITE OUTPUT TAPE 6, 1000
WRITE OUTPUT TAPE, 6, 1001
WRITE OUTPUT TAPE 6, 1002
```

Table VI. Comparison of Computer Speeds

System	Program Development		Production, 1,000 Cases, Machine-Hours	Cost, Dollars		
	Man-Hours	Machine-Hours		Development	Production	Total
Symbolic, *705*	24.0	0.100	0.25	220	75	295
Print, *705*	8,0	0.080	0.28	90	85	175
FORTRAN, *704*	0.2	0.065	0.035	25	12	37
APS III, Datatron *205*	1.0	0.200	1.700	14	50	64

```
      READ 31001, XHT, XKT, XWT
      DO 13 I=1, 12
      DO 13 II=1, 12
      DO 13 III=1, 7
      A=XHT (I)-0.5*XWT (III)
      B=XHT (I)+0.5*XWT (III)
      C=XKT (II)-0.5
      D=XKT (II)+0.5
      HX=0.5*(A*NLOGF ((A*A+C*C)/
        (A*A+D*D)))
    1  +B*NLOGF((B*B+D*D)/
        (B*B+C*C))
    2  +D*(ARTNF(B, D)-ARTNF(A, D))
    3  +C*(ARTNF(A, C)-ARTNF(B, C))
13    WRITE OUTPUT TAPE 6, 32005, XHT (I),
        XKT (II), XWT (III), HX
      END FILE 6
      REWIND 6
      STOP 77777
1000  FORMAT (37H
                   MAGNETIC FIELD PROBLEM
    1        36H                          )
1001  FORMAT (37H
    1        36H                          )
 1002 FORMAT (37H       WT       KT      HT
    1         36H       HX/2PI            )
30000 FORMAT (72H
    1                        )
31001 FORMAT (5E14.0)
32000 FORMAT (1H1)
32005 FORMAT (1H 1P5G14.6)
```

APS III

Programing was done in a symbolic language acceptable to the APS III compiler system for the Datatron *205* as developed by Dr. H. G. Rice, Westinghouse Research Laboratory. The APS III compiler system is similar to the FORTRAN system in that it utilizes a mathematical notation requiring the programer to know very little about the operation of the computer.

Approximately one hour was required to write this problem in compiler language. The total program consisted of 32 lines. No special forms were required for programing, any available paper being satisfactory. The complete program is shown in the following. Upper-case letters in formulas refer to arithmetic and mathematic operations such as sum difference, product, quotient, arctangent, etc. Lower-case letters represent the names of variables.

```
prnt alph
Sharon test problem
x     k/t     h/t     w/t
(A)
form c. r.
eval a=S ht P wt NEG .5
eval b=S ht P wt .5
eval c=D kt .5
eval d=S kt .5
eval x=S S P .5 S P a LOG Q S SQ a SQ c S SQ
        a SQ d P b LOG Q S SQ b SQ d S SQ b SQ
        c P d D ARCT Q b ARCT Q a d P C D ARCT
        Q a c ARCT Q b c
eval x=(PR) x
form tab
tran B pos D (PR) ke kt'
eval kt=S kt .1
tran A
(B)
form tab
eval kt= .1
tran C pos D (PR) ht ht'
eval ht=S ht .1
form c.r.
tran A
(C)
form tab
eval ht= .1
tran D pos (PR) wt wt'
eval wt=S wt .2
tran A
(D)
stop
END
```

The previous instructions or statements are in the compiler language. The compiling routine transformed these 32 lines or statements into 143 machine language instructions (plus 337 instructions for subroutines), the first ten of which are as follows. This, incidentally, is how the program would have to be programed in machine language.

```
40000002000a
00000722005a
10000642006a
00000038410a
00000227021a
00000302015a
00000000008a
02723050500a
02304543043a
00467304327a
```

Compiling and punching the input paper tape required approximately 11 minutes' time. Input to the computer must initially be from paper tape; however, programs could be accumulated on magnetic tape and read into memory similar to the magnetic tape system of the IBM *700* series computers.

The limiting factor for running this problem on the Datatron computer is the typewriter if it is used as output. To illustrate, the magnetic field problem was run both with and without the typewriter. Running time for the problem was 124 seconds for 12 points or 10.3 seconds per point with the typewriter as output. Elimination of typewriter in favor of tape output required 72 seconds for 12 points or 6 seconds per point.

Summarizing, the magnetic field problem required the following approximate times using APS III:

Programing of problem, 1/2 to 1 hour.
Compilation and punching of input tape, 10 minutes.
Running time, 1 minute plus 6 seconds per point.

Comparisons

If actual running time of problem only is considered, the following is an approximate comparison of speed of the various computers for calculation of one point of the magnetic field problem.

IBM *704*, 0.085 second per point
IBM *705*, 0.85 second per point
Datatron *205*, 6 seconds per point

However, the most logical comparison involves cost of program preparation as well as running cost. The superiority of the compilers is shown by Table VI.

◆

Discussion

A. E. Peterson (Bell Telephone Laboratories, Inc., Whippany, N. J.): The authors have accomplished much in pointing out the usefulness of automatic coding systems such as FORTRAN. FORTRAN is a mathematical formula translating system for the IBM *704* computer. As cited in Appendix I, FORTRAN is written in a language close to English and algebra. This feature considerably reduces programing complexities for the noncareer programer, therefore making it attractive to people such as engineers.

It is felt that many readers would do well to read the appendixes first, since they contain a wealth of general information on the subject. Appendix I gives a good general description of several coding systems, including FORTRAN and SAP. Appendix II gives a comparison between four systems, which is nicely summarized in Table I showing cost comparisons. The advantages of automatic coding systems are quite obvious from this table.

I have written programs in both symbolic and automatic coding systems. The automatic system (FORTRAN) was considerably easier to use than the symbolic. A specific instance taken from my memory was a program for floating-point addition. This program, when written in symbolic language, required more than 20 steps, but with FORTRAN, only several statements were necessary.

A Study of Thermal Deterioration of Kraft Pulps Using a Mass Spectrometer

Y. SAITO
MEMBER AIEE

T. HINO
NONMEMBER AIEE

AS THERMAL deterioration of electrical insulating materials is directly connected with their lives it has been an important problem since electric machines and apparatuses were put to practical use. Nevertheless, a thorough study has not been made on the subject because quite a long time is required for testing the lives of the materials and also because numberless combinations and many working conditions of materials must be considered in order to get sufficient data concerning the problem. Moreover new insulating materials are rapidly appearing and it has become necessary for us to know their deterioration properties. Consequently, it is necessary to find a test method which will enable us to know these characteristics within a short period of time. Accordingly the authors intended to find the degradation of some kind of insulating materials with a mass spectrometer as it was impossible to detect a large quantity of deterioration products by the apparatus in general use.[1] Kraft pulps were taken as samples first. In analyzing thermally generated gases from the samples, fundamental data were obtained in a short time by which thermal endurance of the materials could be concluded.

Measurement of Thermal Deterioration with Mass Spectrometer

MASS SPECTROMETER

The mass spectrometer employed in the research was an *RM-A* type manufactured by Hitachi Electric Company. It has a 90-degree type of analyzer tube and a gas container that will hold 10 liters.

SAMPLES

Kraft pulps was taken as samples for the investigation. The pulps having different degrees of beating were prepared. The properties of the samples are summarized in Table I. All samples in the table are of the same origin; the second column shows the degree of beating measured with Schopper-Riegler's apparatus. The "Measurement on Thermal Deterioration" section aims to show how the effect of a slight difference in the treatment of the sample can be detected with a mass spectrometer.

PRETREATMENT OF SAMPLES

Fibrous materials which have lain in open air for a long time have absorbed a considerable amount of water. Consequently all such samples have to be thoroughly dried in a vacuum vessel to expel the moisture in order to get equal initial conditions for all measurements. Kraft pulp is treated in a vacuum vessel at 50 C (degrees centigrade) for 4 hours.

MEASUREMENT OF THERMAL DETERIORATION

A glass flask arrangement, shown in Fig. 1, is used in order to gather the gas driven out from a sample. *A* and *B* are glass flasks, each having a capacity of 300 cc (cubic centimeters). A sample is inserted in *B*. After the flask arrangement has been evacuated and filled with some sort of gas, *B* is heated to a prescribed temperature. Then some of the gas or vapor is produced in *B* and half of the quantity is put into *A*. These processes are all realized by the proper operation of three glass cocks, C_1, C_2, C_3, in Fig. 1 and by a vacuum pump or source of gas which is not seen in the figure. After this *A* is separated from the whole glass arrangement at *F* by fusion. *A* has a double seal by which the gas in *A* is transferred into the glass container of the mass spectrometer by connecting *A* to it and breaking the inner seal. When a sample is to be heated in oxygen or argon a liquid air trap is used to separate the produced gas from these gases. *M* is a McLeod gage and *P* is the connection to the vacuum set. Total pressure of generated gas is measured by *M*, and sometimes it is necessary to correct the indication of *M* when too much water vapor is generated.

AN EXAMPLE OF ANALYSIS

Fig. 2 shows an example of a spectrogram of a sample of Kraft pulp drawn by the mass spectrometer. Table II is the result calculated from the peak values in Fig. 2. The second column of the table shows the height of every peak in Fig. 2 and the third column is the relative value of each gas expressed in molecular per cent compared to H_2O. The result is obtained by heating 10 gm (grams) of Kraft pulp for 4 hours at 135 C. It is clear from the table that H_2O, CO_2, and CO are the main products decomposed from the pulp.

EXPERIMENTAL RESULTS

In each case 10 gm of pulp is used for the following experiments. The result of analysis is given by the pressure of each gas in millimeter (mm) Hg (mercury) in a 1-liter vessel at 0 C, and is generated for 1 hour at each step of temperature. As it is considered reasonable to observe carbon atom in order to know the decomposition of cellulose, CO and CO_2 particularly are picked up.

CHANGE OF PRODUCED GAS WITH LAPSE OF TIME

The change of the quantity of thermally expelled gas from cellulose is examined after a lapse of time and after heating the sample at 135 C. Although the rate of the total gas generation is considerably large in the initial stage of heating it is gradually decreased and settles to a constant rate after prolonged heating. It is rather curious that H_2O does not seem to come out at the initial stage of heating, but is generated at nearly constant rate after some time has elapsed. The ratio between CO and CO_2 fluctuates during the initial period of heating, but it becomes a nearly constant value in from 20–30 hours.

Table I. Properties of Kraft Pulps

Sample*	Degree of Beating	Potassium Permanganate Number	α-Cellulose Content, % More Than	Resin Content, % Less Than	Ash Content, % Less Than
No. 1	Unbeaten	28	85	0.3	1.0
No. 2	58				
No. 3	85				

* The three samples are of the same origin but different in degree of beating.

Paper 59-116, recommended by the AIEE Electrical Insulation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted July 22, 1958; made available for printing April 1, 1959.

Y. SAITO and T. HINO are with The Tokyo Institute of Technology, Tokyo, Japan.

Vol.
Pt.
195

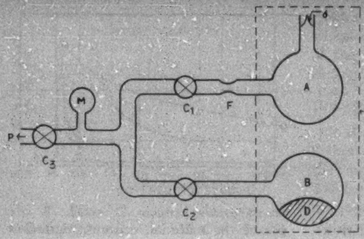


Fig. 1. Schematic diagram of gas-generation apparatus

A = gas captor
B = gas generator
C_1, C_2, C_3 = glass cocks
D = sample
d = double seal
F = portion of fusing
M = McLeod gage
O = oven
P = vacuum pump

Thermal Properties of Kraft Pulp Having Different Degrees of Beating

Fig. 3 compares thermal endurance of samples *1*, *2*, and *3* in Table I. The pretreatment of the samples is continued for 30 hours at 135 C. The abscissa is the reciprocal of treating temperature *T*, and the ordinate is a logarithm of adjusted pressure of $CO+CO_2$ produced from the samples. In Fig. 3 *a*, *b*, *c*, show the relation in oxygen having a pressure of 150 mm Hg; *a'*, *b'*, *c'* are the relations in vacuum, and *a"*, *b"*, *c"* are the relations in argon having 1 atmospheric pressure. All the relations show straight lines having a slight difference in height and inclination, consecutively varied with the change of degree of beating.

Effect of Atmosphere on Thermal Decomposition

Although the difference of thermal deterioration of three samples of Kraft pulp in O_2, vacuum, and argon have been shown in Fig. 3, the properties of sample *1* in oxygen, vacuum, and argon are pointed out again in Fig. 4. For the purpose of pretreatment, the pulp is heated for 4 hours at 50 C. It is easily imaginable that a quantity of generated gas would be large and the ratio CO/CO_2 would be irregular at the beginning of heating even though such a degree of pretreatment has been applied. Nevertheless the relations between $\log P_{CO+CO_2}$ and $1/T$ are exact straight lines having nearly equal inclination as shown in Fig. 4. The quantities of generated gas differ remarkably from one another.

Effect of Beating on Thermal Deterioration

As many samples of Kraft pulp having different degrees of beating have been prepared besides those in Table I, their relation of gas generation and the degree of beating is measured at 115 C in oxygen, keeping the pressure at 150 mm Hg. It is shown in Fig. 5 that the volume of generated $CO+CO_2$ increases as the degree of beating grows stronger. This shows that the thermal deterioration of Kraft pulp is quickened by beating.

Although the change of physical and chemical properties accompanied by the beating has been ascertained already in various aspects,[3] change of the cellulose molecule itself does not seem to occur, i.e., the degree of polymerization, copper number, and the diffraction pattern of X ray do not seem to be altered by beating. Owing to the beating the most remarkable change seems to be that the pulp gets more slender, consequently the effective surface area of the pulp increases and the thermal deterioration is quickened.

Relation Between Mass-Spectrometer Method and Weighing Method

The mass-spectrometer method is compared with the weighing method to see if the data obtained in a short time by the former method are useable as well in judging long-time deterioration. Sample *1* is heated for a very long time at various temperatures arranged for the experiment and weighed every 24 hours to note the decrease of weight. Fig. 6 shows the relations between the time (in days) of heating and weight decrease (in per cent). Anomalous decrease of the weight at the initial stage of heating is not included in the figures. From the results indicated in Fig. 4(A) and those in Fig. 6, the relation in Fig. 7 is derived. This shows the rate of weight decrease of sample *1* pulp for 24 hours when it is heated at each stage of temperature. The three black spots in the figure are obtained actually by weighing method and the remaining points are calculated from the data obtained by mass spectrometer method. The upper limit of the rank of vertical short line shows the value when the oxygen in the generated gas (CO, CO_2) is assumed to have come out from the constituent of cellulose, and the lower limit shows the relation when all the oxygen in the generated gas is assumed to have come from the atmosphere. It can easily be seen from the figure that the straight line obtained by weighing method has nearly the same inclination as those obtained by the mass-spectrometer method and the quantity of generated gas measured by each method seems to coin-

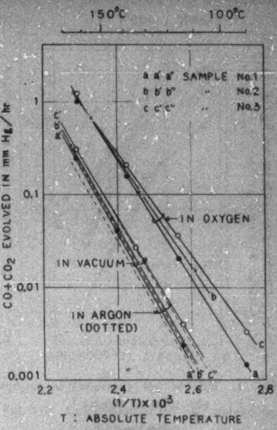


Fig. 3. Rate of gas generation from Kraft pulps in different atmospheres

a, b, c obtained in oxygen
a', b', c' obtained in vacuum
a", b", c" obtained in argon

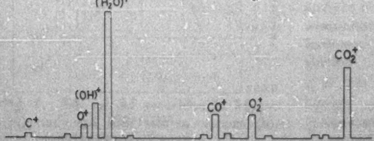


Fig. 2. Spectrogram of gas analysis

Table II. Example of Analysis

Gas	Quantity, Mm	Relative Value, %
H_2O	77.5	100.0
CO_2	43.0	55.0
CO	12.5	15.6
O_2	12.5	15.6

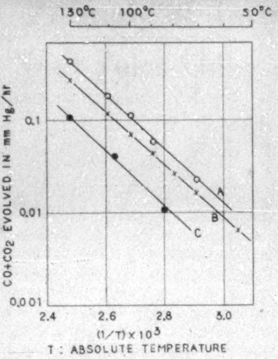


Fig. 4. Rate of gas generation in various atmospheres; sample 1

A—In oxygen
B—In vacuum
C—In argon

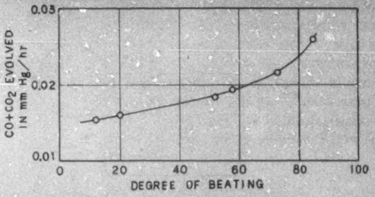


Fig. 5. Relation between rate of gas generation and degree of beating

cide approximately. (Better agreement was ascertained between the results obtained by both methods when samples were tested which had been heated for 20–30 hours at 135 C.) Thus it is possible to assume that the deterioration of the material over a long period of time can be forecasted by the short-time test with a mass spectrometer.

Discussion of Results

Effect of Moisture

The rate of generation of gases from pulp is very large at the initial stage of heating. This may be due to the still remaining moisture in the fibrous structure which causes the generation of a large quantity of $CO+CO_2$. In order to verify such an effect sample *1* is heated for a long time in vaccum at 135 C. When the gas generation has become constant the sample is sufficiently exposed to moisture. Then it is treated in vacuum at 50 C for 4 hours, as formerly mentioned, and heated in vacuum at 135 C again. It is found that the sample shows quite the same character in gas generation as at the first heating. The result is shown in Fig. 8. As the effect of moisture on gas generation from pulp is very serious, every precaution in examining the thermal deterioration of fibrous materials must be taken.

Relation Between $\Delta p \sim 1/T$

The logarithm of the volume of $CO+CO_2$ that is generated for a definite time by thermal decomposition of Kraft pulp has linear relation with the reciprocal of the temperature in absolute scale. This is derived from the theory on chemical reaction as originally carried out by Dakin.[8] The rate of chemical reaction is generally expressed by the following relation:

$$\frac{dc}{dt}=KC^n \quad K=-A\epsilon^{-E/RT} \tag{1}$$

where

C = concentration of material under reaction
E = activation energy
T = absolute temperature
t = time
R = gas constant
A, n = constant

Thermal decomposition of the Kraft pulp is considered as one of the cases represented by equation 1. As the time spent by the thermal decomposition in the present case is very short compared to the life of the material, concentration C is considered to be constant during the reaction. Then the following equation is derived from equation 1:

$$\frac{\Delta C}{C_0^n}=A(\epsilon^{-E/RT})\Delta t \tag{2}$$

where

C_0 = initial concentration
ΔC = decrease of concentration

as CO and CO_2 are generated when cellulose is decomposed, the pressure Δp of $CO+CO_2$ is thought proportional to ΔC, i.e.,

$$\Delta p=\alpha\Delta C \tag{3}$$

Inserting equation 3 in equation 2 and making use of logarithmic expression:

$$\log \Delta p=\log C_0^n+\log A+\log \alpha+\log \Delta t-\frac{E}{RT} \tag{4}$$

As C_0, A, Δt, and n are all constant, equation 4 can be simplified as follows:

$$\log \Delta p=C-\frac{B}{T} \tag{5}$$

where $B=E/R$ and C = a constant.

Equation 5 shows linear relation beween $\log \Delta p$ and $1/T$. This is the same form as the equation representing the relation between the life and the temperature of insulating materials derived first by Dakin and is approved if the generation of CO and CO_2 from the pulp is the main reaction which caused the change of

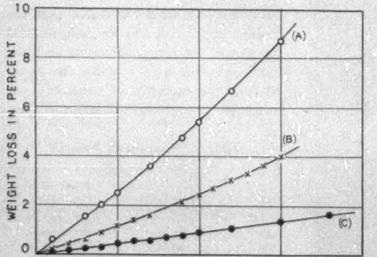


Fig. 6 (left). Deterioration of Kraft pulp tested by weighing method; sample 1

A—170 C
B—155 C
C—130 C

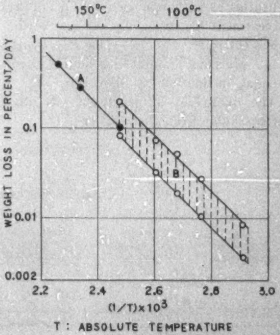


Fig. 7 (right). Mass-spectrometer method compared with weighing method; sample 1

A—Weighing method, Fig. 6
B—Mass spectrometer method, curve (A) in Fig. 4

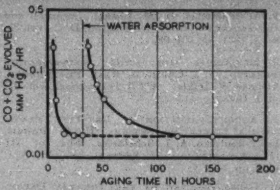


Fig. 8. Effect of moisture absorption on rate of gas generation at 135 C in vacuum, sample 1

physical properties in accordance with thermal deterioration.

Effect of Degree of Beating

The effect of the degree of beating on the thermal deterioration of the Kraft pulp is shown clearly in Fig. 3. As the reciprocal of the rate of the disintegration is considered proportional to the life of the material, it is easily seen from Fig. 3 that the lives of pulp no. *1* and no. *3* in oxygen have the ratio of nearly 1-to-2.5 at 90 C. This fact shows that the thermal endurance of the samples is indeterminable even though the samples have the same origin. As to the straight lines in Fig. 3, their inclinations are distinguishable. The inclination is a little steeper in vacuum or in argon than in oxygen for the same sample. But the straight lines in vacuum and argon have nearly equal inclinations. The less the degree of beating of the sample, the steeper the inclination of the straight line in oxygen. Activation energy of the steepest straight line for sample *1* is 34.6 kilocalorie/molecule in vacuum and that of the least steep line for sample *3* is 24.5 kilocalorie/molecule in oxygen.

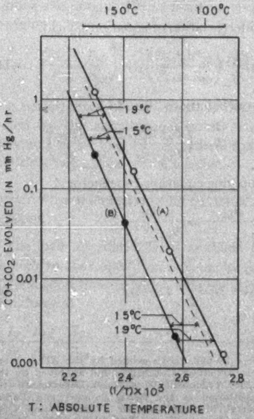


Fig. 9. Rate of gas generation in oxygen (A) compared with rate in vacuum (B), sample 1

Differences Due to Atmosphere

It is clearly understandable from Fig. 3 that the deterioration is promoted by the presence of oxygen as usually expected. The fact that the maximum temperature of class *A* insulating material is 15 C higher than that of class *Y* material is attributed to shutting off oxygen in the atmosphere from the fibrous material by coating it with varnish, oil, etc. Figs. 9(A) and (B) are the same as the straight lines in Fig. 3 and are approximately parallel to each other. From Fig. 9 it is easily seen that if the sample is tested in vacuum at a temperature 19 C higher than in this figure, the straight line expected will coincide with the line in oxygen. This means that if the oxygen atmosphere is thoroughly shut off from the sample and it is put in vacuum it can withstand higher temperature by 19 C than if it were in oxygen. The dotted line in Fig. 9 shows the characteristic at a temperature 15 C lower than the straight line. It corresponds to the state of class *Y* mentioned before.

Comparison of Results on Deterioration Tests

Fig. 10 is the comparison of our results with other authors' results. Figs. 10(A) and (B) are cited from Stewart's report[4] and show the decrease of breakdown voltage of Kraft papers by thermal deterioration in nitrogen and in air. Fig. 10(C) is quoted from Murphy's report[5] which shows the generation of CO+CO$_2$ from Kraft paper by heating. Fig. 10(D) is the result obtained by heating Kraft pulp in vacuum, reported in this paper. The ordinate shows the time required to produce unit volume of CO+CO$_2$ from the pulps at 0 C, 1 atmosphere. Fig. 10(E) is also obtained from the present test in oxygen and shows the same relation as (D). Fig. 10(A) and (C) have nearly equal inclination to the straight line (D)1. Fig. 10 (B) and the straight line (E)3 have nearly equal inclination.

Sensitivity of the Mass Spectrometer

A mass spectrometer enables us to analyze a material as mentioned before. The quartz spring balance we used has a loading capacity of 3 gm and the elongation for the change of the load by 1 gm is 67 mm which is read by a microscope having a magnifing power of 7, so the mass spectrometer is a few hundred times more sensitive than the spring balance. Thus the spectrometer enables us to detect the thermal deterioration of materials in a very short time and even at operating temperature.

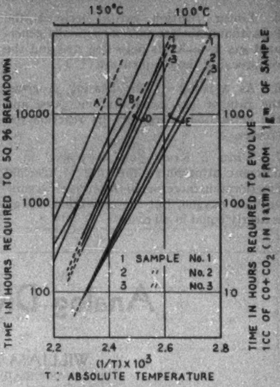


Fig. 10. Comparison of results obtained by different authors

A—Breakdown of Kraft paper in nitrogen, Stewart's data

B—Breakdown of Kraft paper in air, Stewart's data

C—Gas generation of Kraft paper, Murphy's data

D—Gas generation of Kraft pulps in vacuum, authors' results

E—Gas generation of Kraft pulps in oxygen, authors' results

Conclusions

The thermal deterioration of insulating materials is investigated with a mass spectrometer in order to ascertain the merit of the method on the one hand and to get the actual data on Kraft pulps on the other hand. The results obtained are summarized as follows.

1. A mass spectrometer can be effectively applied, for the purpose of detecting thermal deterioration, to those materials which generate such products as gas or vapor by heating. In such a case, thermal deterioration can be tested at working temperature and in a short time.
2. The life of the insulating materials can easily be derived from the aforementioned results.
3. The mechanism of thermal deterioration of insulating materials can be clarified through the temperature range starting from the working temperature and progressing to a considerably high temperature.
4. Applicability of the result obtained at higher temperature for the purpose of judging the life at working temperature is easily noticed from item *3*.

5. Linear relation exists between the logarithm of the pressure of thermally generated gas from Kraft pulp (log Δp) and the reciprocal of the temperature ($1/T$).

6. As a result of the beating progress, the rate of gas generation increases and the activation energy of the reaction decreases.

7. When a Kraft pulp is placed in a different atmosphere the rate of thermal deterioration increases in the order of argon–vacuum–oxygen, but the activation energy is nearly equal in all cases.

8. Water vapor hastens very much the thermal deterioration of the Kraft pulp.

9. Rate of thermal decomposition of unbeaten Kraft pulp in oxygen is equal to that in vacuum at a 19 C higher temperature.

References

1. STUDY OF THERMAL DETERIORATION OF INSULATING MATERIALS USING A MASS SPECTROGRAPH, **Y. Saito, S. Yamanaka, T. Hino.** *Journal, Institute of Electrical Engineers of Japan*, Tokyo, Japan, vol. 77, 1957, pp. 1610–17.

2. EFFECT OF BEATING UPON CERTAIN CHEMICAL AND PHYSICAL PROPERTIES OF PULPS, **C. E. Curran, F. A. Simmonds, H. M. Chang.** *Industrial and Engineering Chemistry*, Washington, D. C., vol. 23, no. 1, 1931, pp. 104–08.

3. ELECTRICAL INSULATION DETERIORATION TREATED AS A CHEMICAL RATE PHENOMENON, **Thomas W. Dakin.** *AIEE Transactions*, vol. 67, pt. I, 1948, pp. 113–22.

4. AGING EVALUATION OF DRY-TYPE TRANSFORMER INSULATING SYSTEMS, **H. C. Stewart, L. C. Whitman, A. L. Scheideler.** *Ibid.*, pt. III (*Power Apparatus and Systems*), vol. 72, Apr. 1953, pp. 267–77.

5. GASES EVOLVED BY THE THERMAL DECOMPOSITION OF PAPER, **E. J. Murphy.** *Transactions-Electrochemical Society*, Baltimore, Md., vol. 42, 1943, pp. 161–74.

Analog Division Circuit

WILLIAM McMURRAY
ASSOCIATE MEMBER AIEE

IN THE FIELD of automatic control, there is a need for simple and reliable devices which perform the operations of analog multiplication and division. This paper describes a circuit containing only magnetic cores and windings, which produces an alternating voltage output directly proportional to the quotient of two d-c input signals. The output is also directly proportional to the alternating voltage used for excitation, which thus may be employed as a multiplier input signal, otherwise a regulated a-c supply is required. Since the circuit contains no semiconductors and is relatively insensitive to changes in the properties of the magnetic cores, it is well-adapted to operation at elevated temperatures.

The circuit, Fig. 1, consists of two similar cross-field reactors A and B, input transformer T_X, output transformer T_L, excitation voltage e_X, d-c control sources $2I_1$ and $2I_2$ in series with impedances Z_1 and Z_2 respectively, and load impedance Z_L. Resistances R_1 and R_2 represent winding resistances. The nonsaturated condition of the transformers maintains equal division of the control currents between the parallel paths at every instant. The turns ratios of the transformers are assumed to be unity in the analysis. An alternative arrangement which eliminates the need for transformers is shown in Fig. 2. It is assumed that the dividend control turns are equal to the excitation turns, and the divisor control turns are equal to the load winding turns.

This circuit is reported in the literature[1] as a multiplying circuit producing an output proportional to the product I_1I_2. However, the conditions under which this action is obtained are not given. The present paper deals with the mode of operation which produces an output proportional to the quotient I_2/I_1, hereafter termed the division mode.

A "hollow doughnut" core shape, Fig. 3, is suitable for construction of the cross-field reactors. A circular winding is inserted in the annular window, and a toroidal winding is put on after assembling the two halves of the core. The magnetic intensity produced by one winding is perpendicular to the magnetic intensity produced by the other winding, throughout the volume of the core. The location of the annular window is such as to maintain approximately uniform core cross-sectional area with respect to the annular winding.

Analysis

It is well known that the action of a cross-field reactor is a result of nonlinearity in the magnetization characteristic of the core. In a previous analysis,[2] this characteristic is expressed as a power series. Considerable simplification is achieved by assuming an isotropic core material having the magnetization characteristic of Fig. 4(A). When both of the mutually perpendicular windings on a core carry current, the magnetic flux density vector $\mathbf{B}_0$ is equal to B_S in the direction of the resultant magnetic intensity vector $\mathbf{H}_0$, see Fig. 4(B).

It follows that*

$$\frac{B_1}{H_1}=\frac{B_2}{H_2}=\frac{B_0}{H_0} \tag{1}$$

$$B_1^2+B_2^2=B_0^2 \tag{2}$$

where

$$B_0=B_S \tag{3}$$

DERIVATION OF GENERAL CIRCUIT EQUATIONS

The winding resistances are assumed to be negligible

$$R_1=R_2=0 \tag{4}$$

With sinusoidal voltage excitation

$$e_X=e_{2A}+e_{2B}=\sqrt{2}\,E_X \sin \omega t \tag{5}$$

$$N_2A_2\frac{d}{dt}(B_{2A}+B_{2B})=\sqrt{2}E_X \sin \omega t \tag{6}$$

Integration of equation 6 yields

$$\omega N_2A_2(B_{2A}+B_{2B})=-\sqrt{2}E_X \cos \omega t \tag{7}$$

Defining the per unit excitation as

$$b_2=\frac{E_X}{2E_{S2}}=\frac{E_X}{\sqrt{2}\omega N_2A_2B_S} \tag{8}$$

Equation 7 becomes

$$\frac{B_{2A}+B_{2B}}{B_S}=-2b_2 \cos \omega t \tag{9}$$

For the load or output circuit

$$e_L=e_{1A}-e_{1B}=N_1A_1\frac{d}{dt}(B_{1A}-B_{1B}) \tag{10}$$

Differentiation of equation 2 yields

$$B_1\frac{dB_1}{dt}+B_2\frac{dB_2}{dt}=0 \tag{11}$$

From equations 10 and 11

$$-e_L=N_1A_1\left(\frac{B_{2A}}{B_{1A}}\frac{dB_{2A}}{dt}-\frac{B_{2B}}{B_{1B}}\frac{dB_{2B}}{dt}\right) \tag{12}$$

Applying equation 1 to reactor A

$$\frac{B_{1A}}{\frac{N_1}{l_1}(I_1-i_L)}=\frac{B_{2H}}{\frac{N_2}{l_2}(i_X+I_2)} \tag{13}$$

$$a(i_X+I_2)=\frac{B_{2A}}{B_{1A}}(I_1-i_L) \tag{14}$$

---

* This method of analysis is due to J. B. McFerran, an associate of the author.

Paper 59-668, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted August 18, 1958; made available for printing April 1, 1959.

WILLIAM McMURRAY is with the General Electric Company, Schenectady, N. Y.

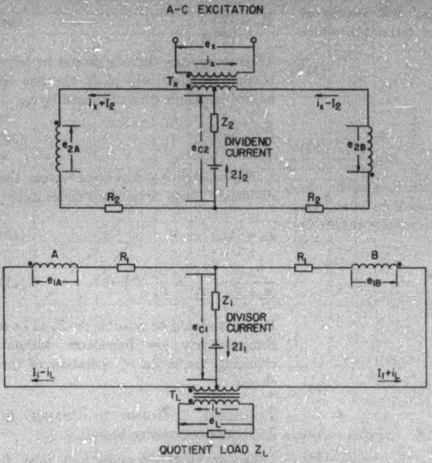

A-C EXCITATION
e_x
i_x
T_x
i_x+I_2
i_x-I_2
Z_2
e_{2A}
e_{C2}
DIVIDEND CURRENT
e_{2B}
$2I_2$
R_2
R_2
A
R_1
R_1
B
e_{1A}
e_{1B}
Z_1
e_{C1}
DIVISOR CURRENT
$2I_1$
I_1-i_L
I_1+i_L
T_L
i_L
e_L
QUOTIENT LOAD Z_L

Fig. 1. Cross-field reactor division circuit

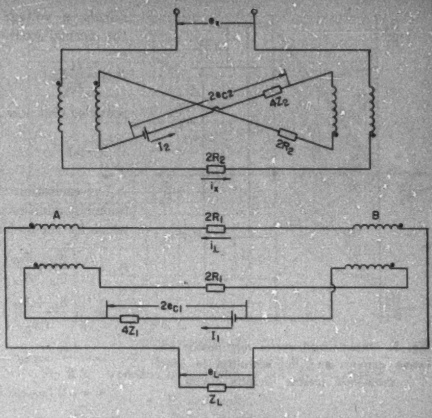

e_x
$2e_{C2}$
$4Z_2$
I_2
$2R_2$
$2R_2$
i_x
A
$2R_1$
i_L
B
$2R_1$
$2e_{C1}$
$4Z_1$
I_1
e_L
Z_L

Fig. 2. Alternative arrangement of cross-field reactor division circuit

where

$$a=\frac{N_2l_1}{N_1l_2}=\frac{N_2A_2}{N_1A_1} \tag{15}$$

since

$$A_1l_1=A_2l_2=\text{core volume} \tag{16}$$

Similarly, applying equation 1 to reactor B

$$a(i_X-I_2)=\frac{B_{2B}}{B_{1B}}(I_1+i_L) \tag{17}$$

Subtraction of equation 17 from equation 14 yields

$$2aI_2=I_1\left(\frac{B_{2A}}{B_{1A}}-\frac{B_{2B}}{B_{1B}}\right)-i_L\left(\frac{B_{2A}}{B_{1A}}+\frac{B_{2B}}{B_{1B}}\right) \tag{18}$$

The performance of the circuit is determined by the simultaneous solution of equations 9, 12, and 18. A general solution does not appear to be feasible.

Steady-State Operation in Division Mode

The load impedance Z_L is assumed to be high so that operation is close to the no-load condition, in which case equation 18 reduces to

$$2a\frac{I_2}{I_1}=\frac{B_{2A}}{B_{1A}}-\frac{B_{2B}}{B_{1B}} \tag{19}$$

If the d-c sources of I_1 and I_2 are steady, and the impedances Z_1 and Z_2 are large, I_2/I_1 is constant.

If it is also assumed that

$$/B_{2A}/\ll B_S \text{ and } /B_{2B}/\ll B_S \tag{20}$$

then

$$/B_{1A}/\approx/B_{1B}/\approx B_S \tag{21}$$

and equation 19 may be written approximately

$$2a\frac{I_2}{I_1}=\frac{B_{2A}-B_{2B}}{B_S} \tag{22}$$

from equations 9 and 22

$$\frac{B_{2A}}{B_S}=a\frac{I_2}{I_1}-b_2\cos\omega t \tag{23}$$

$$\frac{B_{2B}}{B_S}=-\frac{aI_2}{I_1}-b_2\cos\omega t \tag{24}$$

In view of equations 21, 23, and 24, equation 12 may be written approximately

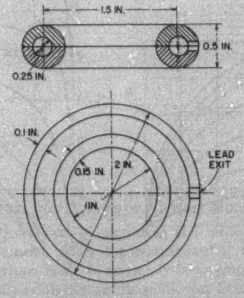

1.5 IN.
0.5 IN.
0.25 IN.
0.1 IN.
0.15 IN.
2 IN.
1 IN.
LEAD EXIT

Fig. 3. "Hollow doughnut" core of cross-field reactor

$$-\frac{e_{LO}}{N_1A_1B_S}=\frac{B_{2A}}{B_S}\frac{d}{dt}\left(\frac{B_{2A}}{B_S}\right)-\frac{B_{2B}}{B_S}\frac{d}{dt}\left(\frac{B_{2B}}{B_S}\right)$$

$$=\left(a\frac{I_2}{I_1}-b_2\cos\omega t\right)\omega b_2\sin\omega t+\left(a\frac{I_2}{I_1}+b_2\cos\omega t\right)\omega b_2\sin\omega t$$

$$=2a\frac{I_2}{I_1}\omega b_2\sin\omega t \tag{25}$$

$$e_{LO}=-\sqrt{2}\,E_{LO}\sin\omega t \tag{26}$$

where

$$E_{LO}=\sqrt{2}\,\omega N_1A_1B_Sab_2\frac{I_2}{I_1} \tag{27}$$

In view of equations 8 and 15, equation 27 reduces to

$$E_{LO}=\frac{I_2}{I_1}E_X \tag{28}$$

Thus the no-load output voltage is equal to the excitation voltage times the quotient of two control currents, and has the same frequency and waveshape as the excitation. (By inference, the latter conclusion also applies if the excitation is nonsinusoidal.) Note that E_{LO} is in-

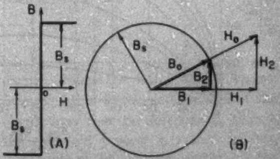

B
B_s
0
H
B_s
(A)
B_s
H_0
B_0
B_2
H_2
B_1
H_1
(B)

Fig. 4. (A) Characteristic of ideal core material. (B) Vector relationship between B and H in ideal cross-field reactor

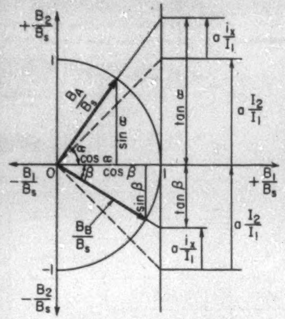


Fig. 5. Normalized vector relationship between currents and flux densities in ideal cross-field reactor division circuit

dependent of the turns ratio N_1/N_2. Multiplication by a constant may be achieved by changing the turns ratio of one or both of the transformers in the circuit of Fig. 1, or between the corresponding reactor windings in the arrangement of Fig. 2.

For inequalities in equation 20 to be valid, it can be seen from equations 23 and 24 that the following restrictions must be met

$$b_2 << 1 \tag{29}$$

$$a\frac{|I_2|}{|I_1|} << 1 \tag{30}$$

Since no restrictions are placed on control current polarities, the circuit is a four-quadrant divider. Reversal of I_1 or I_2 (but not both) will reverse the phase of the load voltage.

The symmetry of the circuit is also noteworthy. If an excitation voltage $\mathbf{E}_X'$ is applied to the normal load terminals, a voltage E_{LO}' will appear at the normal excitation terminals, where

$$E_{LO}' = \frac{I_1}{I_2} E_X' \tag{31}$$

provided that now

$$a\frac{|I_2|}{|I_1|} >> 1 \tag{32}$$

The excitation current i_X may be obtained by addition of equations 14 and 17, which for the no-load condition results in

$$\begin{aligned} i_{XO} &= \frac{I_1}{2a}\left(\frac{B_{2A}}{B_{1A}} + \frac{B_{2B}}{B_{1B}}\right) \\ &\approx \frac{I_1}{2a}\left(\frac{B_{2A}+B_{2B}}{B_S}\right) \\ &= -\frac{I_1 b_2}{a}\cos\omega t \\ &= -\sqrt{2}\, I_{XO}\cos\omega t \end{aligned} \tag{33}$$

where

$$I_{XO} = \frac{I_1 E_X}{2\sqrt{2}a\, E_{S2}} \tag{34}$$

The voltage e_{C1} appearing across the divisor control circuit impedance Z_1 is given by

$$e_{C1} = \frac{1}{2}(e_{1A} + e_{1B}) \tag{35}$$

Through similarity with the derivation of equation 25 from equation 10

$$\begin{aligned} -\frac{2e_{C1}}{N_1A_1B_S} &\approx \frac{B_{2A}}{B_S}\frac{d}{dt}\left(\frac{B_{2A}}{B_S}\right) + \frac{B_{2B}}{B_S}\frac{d}{dt}\left(\frac{B_{2B}}{B_S}\right) \\ &= -2b_2\cos\omega t \times \omega b_2 \sin\omega t \\ &= -\omega b_2^2 \sin 2\omega t \end{aligned} \tag{36}$$

$$e_{C1} = \sqrt{2}\, E_{C1}\sin 2\,\omega t \tag{37}$$

where

$$E_{C1} = \frac{E_X^2}{8aE_{S2}} \tag{38}$$

Thus it appears that Z_1 should be large enough to effectively block the flow of second harmonic current caused by e_{C1}.

$$\frac{E_{C1}}{Z_1} << 2I_1 \tag{39}$$

The voltage e_{C2} developed across the dividend control circuit impedance Z_2 is

$$e_{C2} = \frac{1}{2}(e_{2A} - e_{2B}) \tag{40}$$

$$\frac{2e_{C2}}{N_2A_2B_S} = \frac{d}{dt}\left(\frac{B_{2A}-B_{2B}}{B_S}\right) \approx 0 \tag{41}$$

Hence it should be possible for Z_2 to be a comparatively low impedance without changing the mode of operation of the circuit.

TRANSIENT RESPONSE TO CHANGE IN DIVIDEND CONTROL SIGNAL

Suppose that I_2 changes with time, I_1 remaining constant. In view of equation 22, equation 41 becomes

$$\frac{2e_{C2}}{N_2A_2B_S} = \frac{2a}{I_1}\frac{dI_2}{dt} \tag{42}$$

Thus the dividend control circuit exhibits inductance. In the circuit arrangement of Fig. 1

$$L_2 = \frac{N_2A_2B_S a}{2I_1} \tag{43}$$

For the circuit arrangement of Fig. 2, the inductance is four times the value of equation 43.

The total dividend control circuit resistance (in Fig. 1 when Z_2 is a resistance) is

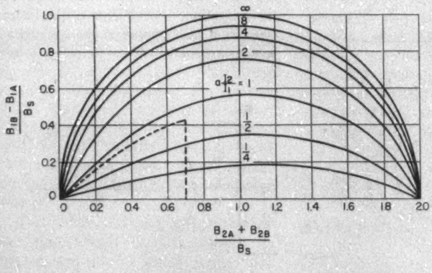


Fig. 6. Relationship between $\frac{B_{1B}-B_{1A}}{B_S}$ and $\frac{B_{2A}+B_{2B}}{B_S}$, with $a\frac{I_2}{I_1}$ as a parameter

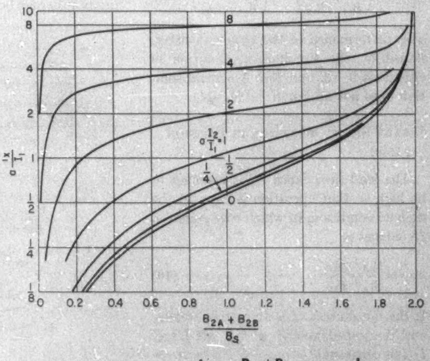


Fig. 7. Relationship between $a\frac{i_X}{I_1}$ and $\frac{B_{2A}+B_{2B}}{B_S}$, with $a\frac{I_2}{I_1}$ as a parameter

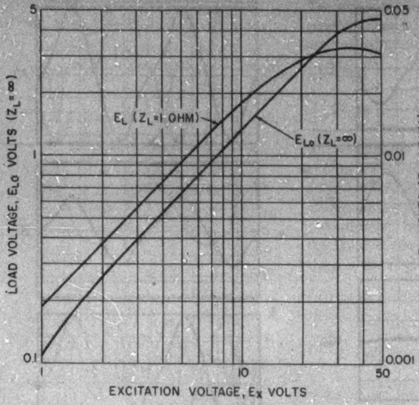


Fig. 8. Variation of load voltage with excitation voltage

$2I_1=0.32$ amp
$2I_2=0.05$ amp

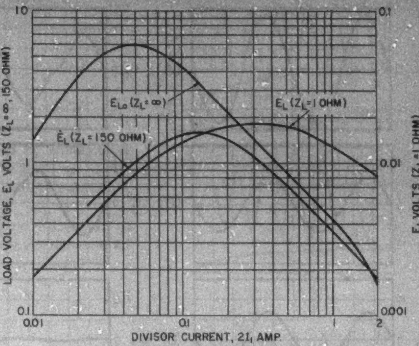


Fig. 10. Variation of load voltage with divisor current

$E_X=10$ volts
$2I_2=0.05$ amp

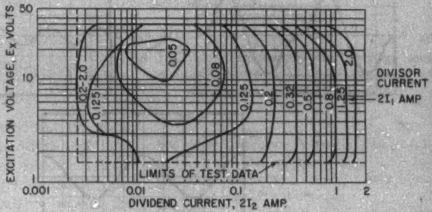


Fig. 11. Contour map showing region within which independent variables E_X, $2I_1$, $2I_2$ must be for $0.7<k<1.0$

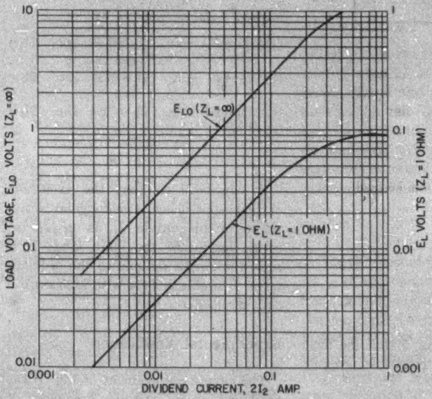


Fig. 9. Variation of load voltage with dividend current

$E_X=10$ volts
$2I_1=0.32$ amp

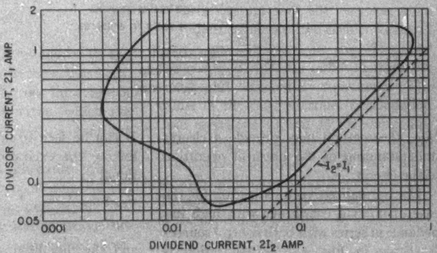


Fig. 12. Region within which $2I_1$ and $2I_2$ must be for $0.8<k<0.9$, when $E_X=16$ volts

$$R_{C2}=\frac{1}{2}R_2+Z_2 \qquad (44)$$

and the dividend control circuit time constant is

$$\tau_2=\frac{L_2}{R_{C2}} \qquad (45)$$

From equations 43 and 45 it is seen that the time of response to a change in dividend control voltage is inversely proportional to the divisor control current.

OPERATION OUTSIDE OF DIVISION MODE

A method of predicting the mode of operation when the restrictions of the division mode do not apply is to graphically relate the net magnetic flux linking the load circuit, the net fluxes linking the control circuits, and the exciting current with the net flux linking the excitation circuit. Since the latter flux is known from equation 9, the performance of the circuit may then be determined, though the complete process is rather tedious. The method is also applicable to the division mode and will aid in estimating the severity of the restrictions necessary to obtain a desired accuracy. In addi-

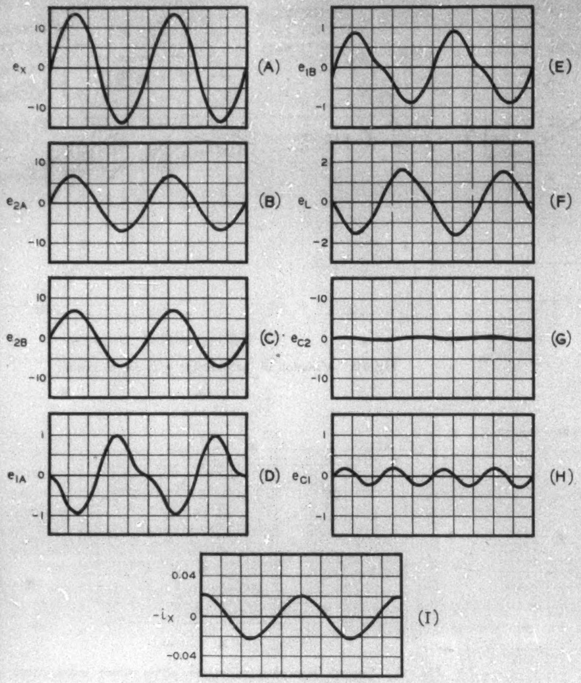


Fig. 13. Oscillograms of steady-state operation in division mode, with no load

Voltage scales in volts
Current scale in amperes

$E_X=10$ volts
$2I_1=0.32$ amp
$2I_2=0.05$ amp

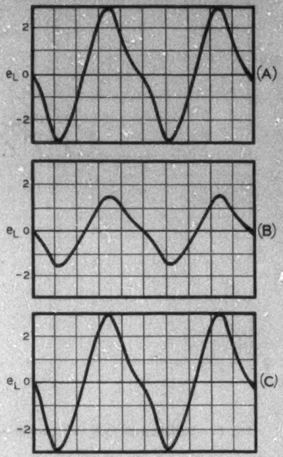


Fig. 14. Oscillograms of load voltage, illustrating validity of equation 28

Voltage scales in volts

(A) $E_X=20$ volts; $2I_1=0.32$ amp; $2I_2=0.05$ amp
(B) $E_X=20$ volts; $2I_1=0.64$ amp; $2I_2=0.05$ amp
(C) $E_X=20$ volts; $2I_1=0.64$ amp; $2I_2=0.10$ amp

tion, the method may provide a clearer understanding of the physical operation of the circuit.

The normalized vector diagram for the case of no-load conditions with high impedance in series with both control sources is shown in Fig. 5. No restriction is placed upon the ratio I_2/I_1. If particular values of a I_2/I_1 and a i_X/I_1 are chosen, then the resultant flux density vector for reactor A makes an angle α with the B_1/B_S axis, where

$$\tan\alpha=a\frac{i_X}{I_1}+a\frac{I_2}{I_1} \qquad (46)$$

The resultant flux density vector for reactor B makes an angle β with the B_1/B_S axis where

$$\tan\beta=a\frac{i_X}{I_1}-a\frac{I_2}{I_1} \qquad (47)$$

It can be seen from Fig. 5 that the flux linking the excitation circuit is proportional to

$$\frac{B_{2A}+B_{2B}}{B_S}=\sin\alpha+\sin\beta \qquad (48)$$

and the flux linking the load circuit is proportional to

$$\frac{B_{1B}-B_{1A}}{B_S}=\cos\beta-\cos\alpha \qquad (49)$$

Equations 48 and 49 are related in Fig. 6, and ai_X/I_1 is related to equation 48 in Fig. 7; the parameter in both cases being aI_2/I_1.

It may be deduced from Fig. 6 that the load voltage will contain a third harmonic component of sufficient magnitude to cause the load voltage to be zero at six instants per cycle when the per unit excitation approaches unity. From Fig. 7 it can be seen that the excitation current approaches a rectangular waveshape of magnitude I_2 as aI_2/I_1 becomes large. In the limit, when I_1 is zero, the circuit reduces to the ordinary series saturable reactor with high control circuit impedance.[3]

Experimental Verification

The circuit arrangement of Fig. 1 was employed, at an excitation frequency of 400 cps. Dimensions of the reactor cores are given in Fig. 3. The material is a General Ceramics Corporation ferrite, "Ferramic H." The annular winding was used for applying the divisor control signal and obtaining the output, while the excitation and the dividend control signal were applied to the toroidal winding. The pertinent parameters are

$B_S\approx 3{,}250$ gausses $=0.21(10^{-3})$ weber/inch^2
$H_C\approx 1$ ampere-turn/inch
$A_1l_1=A_2l_2=0.686$ inch^3
$A_1=0.570$ inch^2; $A_2=0.147$ inch^2
$l_1=1.204$ inches; $l_2=4.66$ inches
$N_1=200$ turns; $N_2=800$ turns
$a=1.034$
$R_1\approx R_2\approx 11$ ohms

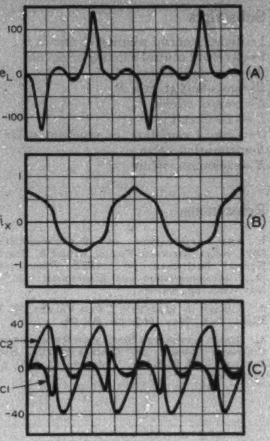


Fig. 15. Oscillograms illustrating operation outside of division mode, with no load

Voltage scales in volts
Current scale in amperes
$E_X=70$ volts
$2I_1=0.25$ amp
$2I_2=1.0$ amp

The impedances Z_1 and Z_2 were chokes of approximately 50 millihenrys.

Figs. 8, 9, and 10 show the results of varying one of the parameters E_X, I_2, and I_1 respectively, while the other two remain constant. The no-load output voltage E_{LO} and the output voltage E_L with resistive load R_L are the dependent variables. When R_L is 1 ohm, the load voltage E_L numerically approximates the load current with the load windings short circuited. With logarithmic scales, a straight line at an angle of +45 degrees shows direct proportionality and a straight line at an angle of −45 degrees shows inverse proportionality. It is observed from Figs. 8 and 9 that multiplication of E_X and I_2 is preserved with loading, while Fig. 10 indicates that division by I_1 is destroyed by excessive loading.

Comprehensive tests were made to determine the accuracy of equation 28. The data are interpreted in terms of an empirical factor k to be used in the equation

$$E_{LO}=kE_X\frac{I_2}{I_1} \tag{50}$$

Over a wide range of the independent variables E_X, I_1 and I_2, the factor k lies within the limits of 0.7–1.0, or a median value of 0.85 with an error of ±17.6%.

The region is mapped in Fig. 11, and the general boundaries may be described approximately as

$$b_2\leq 0.35 \tag{51}$$

$$a\frac{|I_2|}{|I_1|}\leq 0.8 \tag{52}$$

The boundaries defined by equation 51 and 52 are shown by dashed lines in Fig. 6, and verify the analytical predictions of inequalities 29 and 30.

Additional boundaries, attributed to the finite width of the hysteresis loop of the core material, are approximately

$$\frac{N_1I_1}{l_1}\geq 5H_C \tag{53}$$

$$\frac{N_2I_2}{l_2}\geq 0.5H_C \tag{54}$$

An upper limit to the value of I_1 is implied when the winding resistances are neglected, and may be expressed as

$$I_{XO}R_2<<\frac{E_X}{2} \tag{55}$$

which in view of equation 34 becomes

$$I_1R_2<<a\sqrt{2}\,E_{S2} \tag{56}$$

Actually, the reactors become thermally limited when $2I_1$ is approximately 2 amperes. There also appears to be a lower limit to the value of the excitation voltage, attributed to unbalance between the reactors and other stray effects. The excitation current I_{XO} is quite close to the value of equation 34, provided that

$$a\frac{|I_2|}{|I_1|}<0.2 \tag{57}$$

Greater accuracy is possible if the range of the variables is more restricted. For example, Fig. 12 defines the range of the d-c control signals within which the factor k is 0.85±6%, with a constant value of excitation.

Oscillograms which illustrate steady-state operation at a typical point in the division mode are presented in Fig. 13. The waveforms verify the equations derived in the analysis. A sequence, showing the effect upon the load voltage e_{LO} of doubling each of the quantities E_X, I_1, and I_2 in turn, is formed by Fig. 13(F) and Fig. 14(A), (B), and (C). The changes observed confirm the prediction of equation 28.

An example of the type of operation which occurs when the restrictions of equations 51 and 52 are grossly violated is shown in Fig. 15. The conclusions reached by inspection of Figs. 6 and 7, mentioned in the section on operation outside of the division mode, are confirmed.

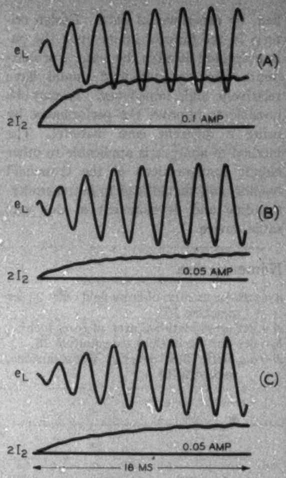


Fig. 16. Response of load voltage and dividend control current to a step input of dividend control voltage

$E_X=10$ volts
(A) $2I_1=0.64$ amp; $R_{C2}=7.5$ ohms; $\tau_2\cong$ 3.2 milliseconds
(B) $2I_1=0.32$ amp; $R_{C2}=15$ ohms; $\tau_2\cong$ 4.0 milliseconds
(C) $2I_1=0.32$ amp; $R_{C2}=7.5$ ohms; $\tau_2\cong$ 6.5 milliseconds

Three examples of the response to a step input applied to the dividend control circuit are presented in Fig. 16. The external choke Z_2 was removed from the circuit and replaced by resistance. The values of the time constant τ_2 calculated from equations 43 and 45 for the conditions of Fig. 16(A), (B), and (C) respectively are 5.35, 5.35, and 10.7 milliseconds, which are in fair agreement with the approximate measured values stated in the caption of Fig. 16.

Conclusions

A pair of cross-field reactors can be employed to perform the operation of analog division. The arrangement has the advantages of simplicity, ruggedness, ability to operate under severe environmental conditions and reasonable accuracy for a wide range of the divisor and dividend control signals. These attributes are important in many applications, such as automatic control systems.

An adequate analysis of the no-load condition has been presented and verified, by experimental data. Design criteria in the form of limits upon the parameters are listed in the text. For maximum accuracy, the load should have relatively high impedance. It may be possible to improve the performance by using a different core material. The method of analysis is applicable to other circuit configurations of the cross-field reactor, and should lead to better understanding and development of this versatile device.

Nomenclature

a = area-turns ratio of cross-field reactor; see equation 15
A = net cross-sectional area of core, inch^2
b_2 = per unit excitation, see equation 18
B = magnetic flux density, instantaneous, webers/inch^2
e = voltage, instantaneous, volts
E = voltage, rms value, volts
H = magnetic intensity, instantaneous, ampere-turns/inch
H_C = coercive force of core material, ampere-turns/inch
i = current, instantaneous, amperes
I = current, a-c rms value or d-c value, amperes
k = empirical factor, see equation 50
l = mean length of magnetic circuit, inches
L_2 = inductance of dividend control circuit, henries
N = number of turns in a reactor winding
R = resistance, ohms
t = time, seconds
Z = impedance, ohms
α, β = angles; see Fig. 5
τ_2 = time constant of dividend control circuit, seconds
ω = angular frequency of excitation, electrical radians/second
d/dt = differential operator with respect to time

Subscripts

1 = divisor control winding or circuit
2 = dividend control winding or circuit
A = cross-field reactor A
B = cross-field reactor B
C = control circuit
L = load circuit
O = no-load condition, or resultant value
S = saturation value
X = excitation circuit

References

1. THE APPLICATION OF PERPENDICULARLY SUPERPOSED MAGNETIC FIELDS, Robert A. Heartz, Herbert Buelteman, Jr. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, Nov. 1955, pp. 655–60.

2. THEORY OF MAGNETIC CROSS VALVES, W. H. Higa. *Ibid.*, May 1955, pp. 204–08.

3. SERIES-CONNECTED SATURABLE REACTOR WITH CONTROL SOURCE OF COMPARATIVELY HIGH IMPEDANCE, H. F. Storm. *Ibid.*, vol. 69, 1950, pp. 1299–1309.

A Transposition System for Carrier Systems Up to 156 Kc

BESS M. KIRKLAND
MEMBER AIEE

THIS PAPER describes a new transposition system designed to fulfill needs in Alabama. The system permits the use of much longer spans than those required by previous transposition designs intended for operation of carrier systems in the range up to 156 kc, and thereby provides for excellent economies in preparing open-wire toll lines for such facilities.

The widespread demand from the public, industry, and the government for more and better communication service has caused an appreciable increase in the use of open-wire carrier systems despite extensive construction of cable and microwave radio relay systems. Dispersal of industry to small towns and the trend toward suburban living have brought a need for toll circuits to areas once considered dormant.

The application of compandored carrier systems has made this circuit growth feasible by open-wire lines. These systems are being manufactured in quantity by many companies, and the inclusion of compandors makes them economically attractive by permitting less stringent construction requirements than those for noncompandored multichannel systems.

Major economies in the use of compandored systems are effected in a transposition system developed by the Bell Laboratories. Known as the *O-1* design, this system is based on the use of 12-inch pin spacing and 4-inch point transposition brackets. The equivalent of 8-16-8 inch pin spacing is obtained by the installation of pinch brackets at specified locations. A 2-span transposition interval based on nominal pole spacing of 155 feet establishes the maximum lengths for the transposition sections. Some existing lines on which compandored systems could be used to excellent advantage have, however, an average span length which exceeds that of the *O-1* design.

The Montgomery-Smithville Toll Line, from Montgomery to Eufaula, Ala., is a line of this type. An expected increase from 41 circuits in 1957 to 85 circuits in 1965 required the installation of two *OB*, four *OC*, and three *OD* systems in 1958. The transposition system on this 2-arm route was not suitable for systems in the *OC* and *OD* range, and *OB* systems were in service on the four assignments suitable for their use. Seventy miles of the line had a nominal span in excess of 200 feet, while only 7 miles were spaced to conform with *0-1* requirements. Reconstruction of this line, which had a maximum of 20 toll wires, all 104 copper Standard Wire Gage no. 12, necessitated the replacement of 35% of the poles and two pairs of wires.

Three plans for solving the transposition problem were considered. The first consisted of reconstructing 70 miles of line to meet *0-1* spacing requirements by relocating approximately half of the poles to be replaced and increasing by 45% the number of poles and crossarms in the route. Adapting the *0-1* design to a 1-span interval was the second

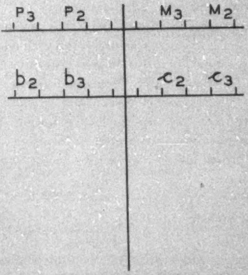


Fig. 1. Pole head diagram of longest type section

Paper 59-509, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Southeast and South Central District Meeting, Atlanta, Ga., April 8–10, 1959. Manuscript submitted February 2, 1959; made available for printing May 28, 1959.

BESS M. KIRKLAND is with the Southern Bell Telephone and Telegraph Company, Birmingham, Ala.

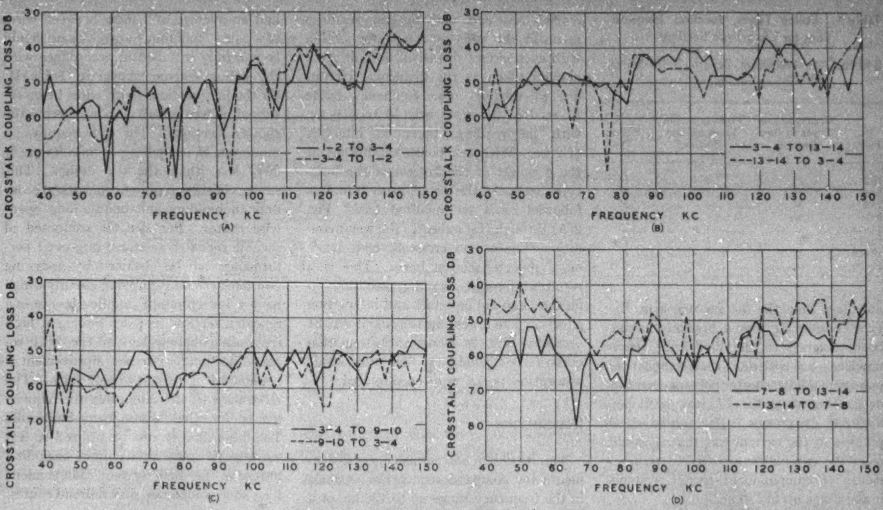


Fig. 2. Far-end equal-level crosstalk coupling for a repeater section. A—Horizontally adjacent pairs. B—Vertically adjacent pairs. C—Alternate pairs. D—Diagonal pairs

plan considered, but the use of 8-inch brackets, to provide 8-16-8 inch pin spacing, with spans more than 200 feet in length would result in excessive wire slapping[1] for the type service required. The third plan was to design a new transposition system for this type of route. Because use of the *0-1* design under these circumstances was economically undesirable, and because its adaptation would have impaired transmission, the design of the new system was undertaken.

Such a system must provide crosstalk coupling loss of at least 30 decibels between pairs to insure satisfactory performance of like compandored carrier systems in the frequency range up to 156 kc. The transposition interval for crosstalk control in this frequency range necessitated a 1-span interval if the existing spacing were to be retained. This shortened the section lengths sufficiently to permit the use of 12-inch point-side brackets instead of the 4-inch transposition brackets and pinch brackets employed in the *0-1* design. The selection of suitable transposition types[2] to control systematic crosstalk at the frequencies involved was extremely critical with the near-end and far-end coefficients resulting from 12-inch pin spacing. Poling types for junction transposition designs were selected carefully to reduce type unbalance crosstalk between all pair combinations.[3] The pole spacing unbalance of three times the section length[1] (*3L*) was achieved largely by relocating deteriorated poles without increasing appreciably the number of poles in the line. Maximum wire sag difference for wires of a pair was established at 3 inches with an average of 1 inch. These limitations were specified to control adequately the components of random irregularity crosstalk.

Fig. 1 shows the fundamental types chosen for the longest section of the system, while Table I lists the types for two arms for all the sections. Computations and experience have shown the types chosen to be satisfactory for noncompandored systems using staggered frequencies, or compandored systems without staggering advantage, in the range up to 40 kc The hardware required for this design is approximately the same as that needed for the *0-1* for two arms, but consider-

Table I. Fundamental Types for Sections

B1 Transposition System

Section Type	Pairs 1-2	3-4	7-8	9-10	11-12	13-14	17-18	19-20	Maximum Length, Miles
B1J	P_3	P_2	M_3	M_2	b_2	b_3	c_2	c_3	5.38
B1K	p_1	P_1	o_1	O_1	B_1	b_1	A_1	a_1	2.69
B1L	*D*	*c*	*E*	*f*	*d*	*C*	*e*	*F*	1.34
B1M	*D*	*E*	*C*	*F*	*H*	*A*	*G*	*B*	0.67
B1N	*I*	*M*	*N*	*J*	*N*	*J*	*I*	*M*	0.33
B1O	*O*	*M*	*O*	*M*	*M*	*O*	*M*	*O*	0.16

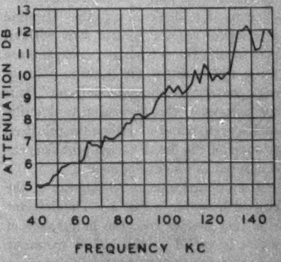


Fig. 3. Attenuation curve of 104 copper pair 12-inch spacing for 34-mile repeater section

Table II. Poling Types, Junction Transpositions for Long-Type Sections

B1J Section Maximum Length 5.38 Miles, 8-Section Poling

Pair	Transposition Types in One Section	Polarity in One Transposition Section	Poling Type
1-2	P_3	++	*P*
3-4	P_3	++	*P*
7-8	M_3	+−	*P*
9-10	M_3	+−	*K*
11-12	b_3	++	*L*
13-14	b_3	++	*P*
17-18	c_3	+−	*P*
19-20	c_3	+−	*L*

ably less than the *0-1* for one arm, the types employing fewest brackets having been assigned to the first arm. Inasmuch as a line with more than eight carrier pairs would have shorter pole spacing, this design, known as the *B1*, was developed for only two arms. The fundamental premise of the system was that it would meet the crosstalk coupling loss requirements of compandored carrier systems in the range up to 156 kc.

Extensive crosstalk computations, combined on a repeater-section basis, were made prior to the installation of the system, and indicated satisfactory performance. After the retransposing was completed, more than 11,000 tests were made on the circuits. These tests included near-end and far-end readings made at the ends of a repeater section with each circuit disturbed and all other circuits taken as the disturbing circuit, one at a time. The frequency range covered in the tests was 40 to 150 kc, with the readings taken at intervals of 2 kc. With test points established at the *S* points of the longest of the long-type sections, the same procedure was followed on a sectionalized basis. Fig. 2(A) through (D) indicates the worst conditions for far-end crosstalk tests made on a repeater-section basis. This is a resultant far-end coupling including reflected near-end crosstalk and interaction effects. The lowest measurement was 34 decibels. Fig. 3 shows an attenuation curve for a typical pair. No undesirable absorption peaks were encountered.

Conclusions

The *B1* design will furnish eight assignments for compandored carrier systems in the frequency range up to 156 kc, or a total of 130 circuits if the pole pairs are used. A maximum average span of 220 feet can be utilized with 12-inch pin spacing and 12-inch point-side brackets. Pole-spacing deviation of three times the section length is permissible, with a maximum wire sag difference of 3 inches, and an average of 1 inch, between wires of a pair. Junction transpositions should be carefully installed in accordance with the poling scheme, shown in Table II, for the long sections, with wires being returned to zero polarity upon entering the shorter sections. The *B1* requires a minimum of ten poles per mile fewer, or 29% less, than the *0-1* design. This savings in poles represents a marked economy in building and maintaining open-wire routes. For systems composed of several repeater sections, improved performance can be obtained by assigning two like systems occupying circuits which have a low crosstalk coupling loss in one repeater section to pairs having a high crosstalk coupling loss in the next repeater section.[1] Further improvement is possible by omitting the pole pairs. The *B1* transposition system met the requirements for the Montgomery-Smithville Toll Line, and it can be utilized to advantage on many similar lines, including routes of the Bell System, independent telephone companies, and railroad routes.

References

1. Open Wire Crosstalk, A. G. Chapman. *Bell System Technical Publication*, New York, N. Y., Monograph B-783, pp. 43, 60, 61.

2. Crosstalk on Open Wire Lines, W. C. Babcock, E. Rentrop, C. S. Thaeler. *Ibid.*, Monograph 2520, pp. 103, 158.

A Rod-Position Indication System for Pressurized Reactors

R. C. FLOYD
MEMBER AIEE

J. F. REUTHER
MEMBER AIEE

THE MOST practical present-day systems for determining the absolute position of reactor control rods without penetrating the pressurized enclosure are those which detect the position of magnetic material in the rod extension. It is the purpose of this discussion to describe such a system and to outline the solutions to some of the detail design problems encountered in its development.

A simplified schematic diagram of the system is shown in Fig. 1. The detector is a long slender coil, the inductance of which is varied as the hollow magnetic rod extension passes over it. The change of inductance is a measure of rod position and is used to light a limit light and operate a position indicator. The detector is in one leg of an Owen bridge. The output of the bridge is rectified by a transistorized phase-sensitive demodulator and fed into the control winding of a magnetic amplifier (preamplifier). This preamplifier is driven from cutoff to saturation in a relatively small portion of the rod travel and its output is fed into the next stage of magnetic amplification (lamp amplifier). The lamp amplifier turns on a lamp that indicates to the operator that the rod is close to its bottom position. The position at which the lamp comes on is adjustable by controlling the bias on the preamplifier using potentiometer *P*1. Bistable magnetic amplifiers were not used since experimentally they were very sluggish with the large power gain required. A selector switch (*S*1) is used to switch a milliameter to read the line voltage by connecting it to the bias supply through suitable resistors or to read the position of any detector by connecting it to the output of the phase-sensitive demodulator. All magnetic amplifiers and lamps are de-energized when reading a rod position since the lamps are out of calibration at this time. A continuously adjustable autotransformer is used to adjust the supply voltage before reading.

Design Considerations

The main design objective was to obtain an output from the unbalanced bridge which varied linearly with rod position. The major design problems resulted from unwanted variables dependent on voltage, temperature, or frequency. Specified

Paper 58-1319, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted August 25, 1958; made available for printing January 8, 1959.

R. C. Floyd and J. F. Reuther are with Westinghouse Electric Corporation, East Pittsburgh, Pa.

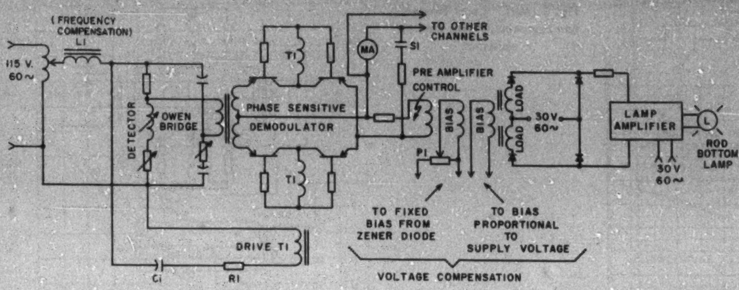


Fig. 1. Simplified schematic of rod position indication system

operating conditions included: line voltage regulation of ±10%, frequency $\pm 4^1/_4\%$, ambient temperature of detector from 40 F (degrees Fahrenheit) to 350 F, and ambient temperature of receiver from 40 F to 149 F. In addition to selecting accurate and stable components, it was necessary to compensate for the following variables:

1. Detector coil resistance with temperature.
2. Detector coil inductance with temperature.
3. Detector coil inductance with voltage.
4. Bridge output with frequency.
5. Meter resistance with temperature.
6. Capacitors with temperature.

Balanced Owen Bridge

Before considering the various compensations in detail, it is well to review the characteristics of the Owen bridge since all measuring system components are essentially part of the bridge.

The Owen bridge is shown in Fig. 2. L_u is the general unknown inductance while L_{uo} refers to the unknown inductance under balance conditions. The balance equations are available in handbooks and are

$$L_{uo}=ASC_B \tag{1}$$

$$U=(C_B/C_A)S \tag{2}$$

No literature was available on the operation of the bridge under unbalanced conditions. A mathematical investigation showed C_A should equal C_B for maximum bridge output voltage. Also, it is desirable to have the bridge internal impedance close to minimum. For a particular application, a minimum impedance of 130 ohms was obtainable for $C_A/C_B=3$ but using $C_A/C_B=1$ only increased the impedance to 150 ohms. It was therefore decided to let $C_A=C_B$.

The value used for C_A and C_B was called C while the value used for U and S was called R. Then, the bridge internal (looking in) impedance at balance is

$$Z_{34}=(R+1/jwC)(1+jwL_{uo}/R)/(2+jwL_{uo}/R) \tag{3}$$

Also the bridge impedance (load seen by the source voltage E_{in}) at balance is

$$Z_{12}=(2R+jwL_{uo})/(1+jwRC) \tag{4}$$

Unbalanced Owen Bridge

INDUCTANCE UNBALANCE

If the detector inductance increases from L_{uo}(value at balance) to L_u, then the open circuit output voltage is:

$$E_{out}=E_{in}(jw)(L_u-L_{uo})/[4R-\omega^2L_uL_{uo}/R+2jw(L_u+L_{uo})]$$

For a particular case the magnitude of E_{out} can be approximated within 1% over the range as:

$$|E_{out}| \approx \text{(approximate equals)}\ |E_{in}|\omega(L_u-L_{uo})/(4R)(1.025) \tag{5}$$

Note that the output varies with the change of inductance.

FREQUENCY DEPENDENCE

While the bridge is theoretically insensitive to frequency at balance, its output under unbalanced conditions is approximately directly proportional to frequency (see equation 5).

RESISTANCE UNBALANCE

If the bridge is balanced and then resistor U is changed (detector coil resistance varies with temperature), the output will be as follows:

$$E_{out}=E_{in}(U/R-1)/(2+2U/R-\omega^2L_{uo}^2/R^2+(3+U/R)jwL_{uo}/R) \tag{6}$$

The denominator stays reasonably constant. The output voltage then varies directly with the error in U. In a particular application, a change of 0.5% in U gave an output equivalent to 3.7% of full rod travel.

CAPACITANCE UNBALANCE

If the bridge is balanced and both capacitors are changed slightly (due to ambient temperature change) the bridge will have an output, as shown by the following:

$$E_{out}=E_{in}jwL_{uo}(1-C/C_o)/R[4-(C/C_o)(\omega^2L_{uo}^2/R^2)+2jwL_{uo}/R(1+C/C_o)] \tag{7}$$

C_o is the value of C at balance. The denominator does not vary much with C, so, approximately:

$$|E_{out}|=|E_{in}|4R(1-C/C_o)/\omega L_{uo} \tag{8}$$

Quantitatively for a specific case, a change in C of 1.7% results in an output of 3.7% of full rod travel.

Compensation

Returning now to specific compensation methods, it is natural to start with the resistance variation of the copper

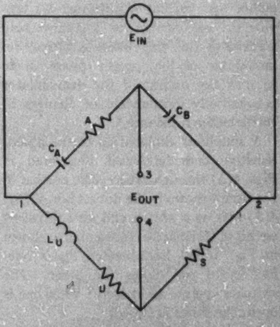


Fig. 2. Owen bridge

Balance equations:

1—$L_{uo}=ASC_B$

2—$U=\frac{C_B}{C_A}S$

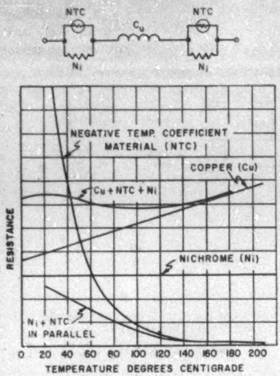


Fig. 3 (left). Local temperature compensation for detector resistance

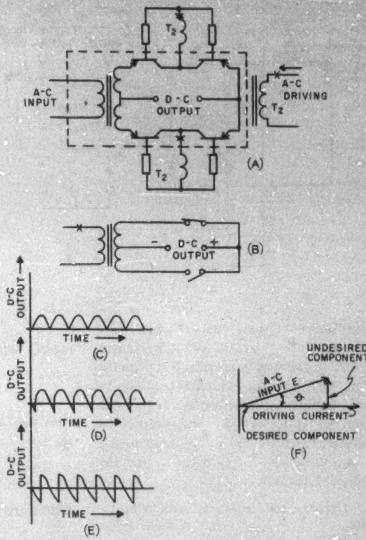


Fig. 4 (right). Phase sensitive demodulator operation

A—Basic circuit
B—Equivalent circuit
C—Driving current in phase
D—Driving current lagging
E—Driving current 90 degrees out of phase
F—Output diagram

detector winding over the wide range of operating temperature (4 C (degrees centigrade) to 177 C), resulting in a change of about 68%. This was corrected to approximately ±6% by local compensation with Nichrome* wire and thermistors, and to less than ±0.25% by phase-sensitive circuitry at the receiver. The method of local compensation is illustrated in Fig. 3, which shows resultant leveling achieved by summing three characteristic curves of resistance versus temperature.

The remote compensation for detector resistance change with temperature is based on the fact that the output voltage of the Owen bridge due to resistance unbalance is approximately 90 degrees out of phase with the output of the bridge due to inductance unbalance. Referring to Fig. 1, condenser *C*1 and resistor *R*1 together with current transformer *T*1 provide driving current to the switching transistors in the phase-sensitive demodulator of the proper phase angle so that the output of the demodulator is essentially independent of changes in the detector resistance.[1]

A simplified explanation of the phase-sensitive demodulator is illustrated in Figs. 4(A) through 4(F). The portion of the circuit shown in the dotted box of Fig. 4(A) acts as a switch and an equivalent circuit of this may be represented as shown in Fig. 4(B). Each switch opens one-half cycle and closes the other. The switches operate alternately; when one is open, the other is closed.

Assume that the top switch closes as soon as the a-c input changes polarity so that the terminal with the polarity *X* is positive. The d-c output has the polarity shown. When the a-c input reverses polarity, assume that the top switch opens and the bottom switch closes. The d-c output again has the same polarity and this amounts to a synchronous rectifier. The output is shown in Fig. 4(C).

Now assume that switches act at some time later than current zero and actually do their switching at an angle θ. Since the switch can carry current in both directions, the output will then appear as shown in Fig. 4(D). An extreme condition would be $\theta = 90$ degrees, in which case the output would look like Fig. 4(E). Note that the net d-c is zero. If θ is greater than 90 degrees but less than 270 degrees, the output reverses polarity.

The average (d-c) value of the output if the input is $E_m \sin wt$ is

$$E_{dc} = (1/\pi)\int_{\theta}^{\pi+\theta} E_m \sin (wt) d(wt) \tag{9}$$

$$= -(E_m/\pi) \cos (wt)\Big|_{\theta}^{\pi+\theta} = (2E_m/\pi) \cos \theta$$

$$= E_{rms} \text{ a-c input } (\cos \theta)/1.11 \tag{10}$$

The d-c output is proportional to the phase angle between the a-c input and switching action. If $\theta = 90$ degrees, the output is zero since cos 90 degrees = 0.

The switching transistor circuit acts similarly. During one half-cycle, one pair of transistors are driven to saturation and act as a closed switch while the other pair are cut off and act as an open switch. During the next half-cycle their action is reversed.

The angle θ used to describe the action of the synchronous switches is now the angle between the a-c input voltage and the a-c driving current. The same cosine law applies.

The output of the Owen bridge under unbalanced conditions consists approximately of two components. One is due to unbalance of inductance while the other is due to the unbalance of the resistance in the inductive leg. These two components are approximately 90 degrees out of phase. The first component is used for metering while the second introduces undesired errors.

If the transistor driving voltage is in phase with the desired component, the effect of the undesired component will be minimized. Referring to Fig. 1, resistor *R*1 and condenser *C*1 are chosen to give the desired transistor driving current at the desired phase angle. The resistors in series with the transistor bases help divide the driving current equally between transistors.

Using the phase angle of the desired component as a reference, the bridge output may be shown as in Fig. 4(F). The desired component is $E \cos \theta$. The output of the bridge E is the a-c input to the phase-sensitive demodulator. The phase-sensitive demodulator output is ($E \cos \theta/1.11$) where E is in rms volts and the output is in d-c volts.

It is apparent from this that, in addition to the fact that the phase-sensitive

---

* Trade name of Driver-Harris Company, Harrison, N. J.

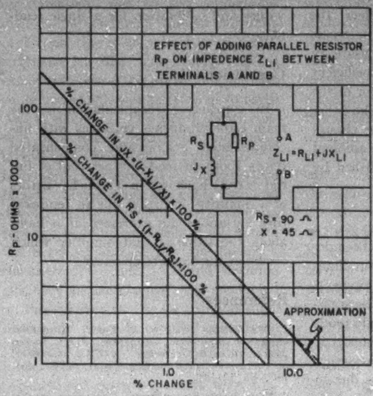


Fig. 5 (left). Shunting resistor effect

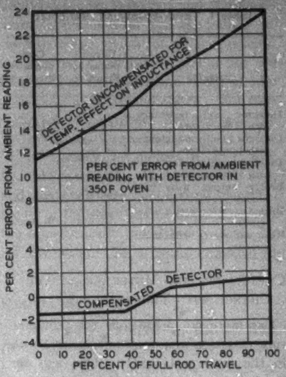


Fig. 6 (right). Compensation for inductance change with temperature

demodulator makes it practical to rectify the low-voltage output of the Owen bridge, it also minimizes the errors attributable to changes of resistance in the detector inductance winding caused by changes in temperature.

Variations of Detector Inductance with Temperature

Temperature tests indicated that the inductance of the detector varied with temperature, approximately 4% increase at each rod position linearly from 90 F to 350 F. Compensation for this variable was approached on the basis that: 1. addition of a resistor (R_p) across the detector-coil terminals will decrease the reactance seen by the bridge, and 2. a thermistor-resistor combination used for the parallel resistor could possibly vary with temperature in such a way as to automatically change the effective reactance to compensate for the change of detector-coil inductance with temperature. Where R_s is the detector resistance, the compensating circuit was developed as follows:

$$Z_{L1} = R_p(R_s + jX)/(R_p + R_s + jX) \quad (11)$$

$$Z_{L1} = (R_p^2 R_s + R_p R_s^2 + R_p X^2)/(X^2 + (R_p + R_s)^2) + jXR_p^2/(X^2 + (R_p + R_s)^2) \quad (12)$$

where X is the detector reactance and Z_{L1} is the impedance of the compensated detector, as shown in Fig. 5.

The real term of equation 12 shows that the equivalent series resistance has changed; however, the phase-sensitive demodulator is insensitive to small changes of resistance. The imaginary term shows that adding R_p reduces the effective reactance by $R_p^2/(X^2 + (R_p + R_s)^2)$. The correction in X is $1 - R_p^2/(X^2 + (R_p + R_s)^2)$. For corrections of about 5%, X is negligible in the denominator (in the range of R and X encountered) and the correction becomes:

$$1 - R_p^2/(R_p + R_s)^2 = 1 - 1/(1 + R_s/R_p)^2 = 2R_s/R_p - 3(R_s/R_p)^2 + 4(R_s/R_p)^3 \ldots \quad (13)$$

Only the first or first two terms of equation 13 are significant. The correction factor versus R_p plotted on log-log paper is a straight line according to Fig. 5. Even the first term $2R_s/R_p$ is only off 0.15% at 4% correction so that a good approximation for the correction is $2R_s/R_p$. The change in effective series resistance was also calculated to ensure that it would not be excessive. From these calculations it was indicated that the correction factor for resistance is approximately $3R_s/4R_p$ compared with $2R_s/R_p$ for reactance. This correction factor for resistance is also plotted in Fig. 5, and the relationship between the two factors is approximately:

$$(\%\text{ change of } R) \approx 3/8\ (\%\text{ change of } X) \quad (14)$$

The characteristics of the thermistor required were taken from data thus calculated and a set of curves showing effect of the compensation for inductance variation with temperature on a production channel is shown in Fig. 6.

Variation of Inductance with Bridge Supply Voltage

Tests showed that if the Owen bridge was balanced at one voltage it would not be balanced at another input voltage level. This was apparently due to variation of the detector inductance with ampere turns, and compensation was desirable. No problem would result when metering since adjustment of voltage with the control provided was a procedural requirement when reading rod position on the test channels. However, the rod bottom-indicator circuitry must be operative under all voltage conditions; hence, compensation for voltage variations was accomplished by providing two biases for the amplifiers, one fixed from a Zener diode and one that varies with supply voltage.

Frequency Compensation

Frequency compensation is necessary since the unbalanced Owen bridge is sensitive to frequency (see equation 3). The output voltage is proportional to both ω and the input voltage. If the product is held constant the bridge would be frequency insensitive. This was approximated by adding a choke in series with the bridge and using a higher supply voltage. Calculation showed that about 1.3 henrys would be optimum in this case, as indicated by the plot shown in Fig. 7.

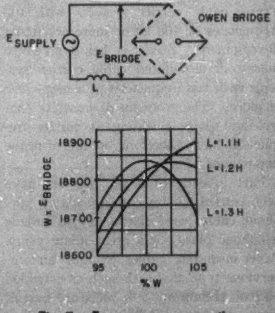


Fig. 7. Frequency compensation

Table I. Error-Producing Variables

Variable	Uncompensated Error, %	Compensated Error, %
Detector coil resistance with temperature	±250	±0.25
Detector coil inductance with temperature	± 5	±1.5
Detector coil inductance with voltage	± 6	±0.9
Bridge output with frequency	± 5	±0.35
Bridge capacitors with temperature	± 2.7	±0.25
Read-out meter with temperature	± 1.5	±0.15

CAPACITANCE AND METER RESISTANCE CHANGE WITH TEMPERATURES

Two more compensations for temperature variation were found necessary, both in the receiver circuitry. One was for the bridge capacitors, which were found on test to vary with temperature more than anticipated. The solution to this unwanted variable was to select the capacitors with a negative characteristic within certain limits and compensate by using a positive temperature coefficient resistor in series with the resistor in the capacitive leg of the bridge. The other compensation was for resistance change in the meter over the rather wide ambient specified, and this was accomplished by the thermistor method.

GENERAL

The error compensations have been described on the basis of independent channels of rod position indication equipment. It may be noted from Fig. 1 that the application discussed employed a single read-out instrument with means to transfer it to multiple channels. Overall accuracy of the system was somewhat less due to the necessity of matching channels with resultant compromise of compensating values and calibration for a single read-out.

Conclusions

In conclusion, recapitulation of the main error-producing variables with uncompensated errors and compensated errors (both in read-out per cent of full rod travel) is shown in Table I. Errors in a quantity occurring due to temperature variations are expressed as a band about the value of that quantity which occurs at the mean temperature.

References

1. TRANSISTOR DEMODULATOR FOR HIGH-PERFORMANCE MAGNETIC AMPLIFIERS IN A-C SERVO APPLICATIONS, **R. O. Decker.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, Mar. 1955, pp. 121–23.

2. THE RADIO ENGINEERING HANDBOOK (book), **Keith Henney**, Editor-in-chief. McGraw-Hill Book Company, Inc., New York, N. Y., 1935, pp. 226–27.

Transmission Properties of Polyethylene-Insulated Telephone Cables at Voice and Carrier Frequencies

G. S. EAGER, JR.
MEMBER AIEE

L. JACHIMOWICZ
MEMBER AIEE

I. KOLODNY
ASSOCIATE MEMBER AIEE

D. E. ROBINSON
ASSOCIATE MEMBER AIEE

IN THE LAST few years, amplitude-modulated carrier telephone systems up to about 250 kc have been routed over polyethylene toll cables and frequency-modulated carrier telephone systems up to about 350 kc have been routed over entrance cable for open-wire circuits. Trunk cable carrier systems with top frequencies of 400 kc or more are under development. Study is being given to pulse-code-modulated cable carrier systems with top frequencies of a megacycle or more. The problems of crosstalk control and noise shielding at carrier frequencies have, therefore, become of prime importance.

Past investigations of this sort have been carried out for paper-insulated telephone cables used extensively in outside telephone plant for a great many years. Paper-insulated cables require an impervious type of sheath such as extruded lead or aluminum or a moisture barrier such as Stalpeth.[1]

In the past decade plastics have become of interest for telephone cables with development centered around polyethylene because of its low dielectric constant, low power factor, high dielectric strength, and electrical stability over a wide frequency range.

In present-day constructions, polyethylene-insulated telephone cables are used in 19, 22, 24, and 26 Awg (American wire gage) copper conductors, and cable sizes up to 900 pairs. They are engineered to have the same mutual capacitance as the standard types of paper cables in order to allow interchangeability and the use of standard loading systems. Cables intended for exchange area use are nonquadded and are designed for average mutual capacitance of 0.083 μf (microfarad) per mile and toll grade nonquadded cables are designed for average mutual capacitance of 0.066 μf per mile.

Until quite recently, polyethylene-insulated telephone cables have been used mostly for the transmission of voice frequency. A cable design to control crosstalk between circuits properly has offered little difficulty. Also, the problem of shielding the cable pairs from external fields at carrier frequencies has seldom been encountered.

Comparatively little transmission data have been published on the types of thermoplastic telephone cable in current use. Attenuation data[2] up to 300 kc are available on older types of cables. It is the purpose of this paper to present a study of the transmission and crosstalk properties of standard, paired, polyethylene-insulated, fully color coded, and Alpeth-sheathed cables throughout the audio- and carrier-frequency range up to 1,000 kc including the effects of temperature variation and moisture on cable parameters. A study of shielding in the same frequency range is also made.

Cables Tested

Most measurements on parameters and all measurements on crosstalk were made

---

Paper 59-778, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing April 9, 1959.

G. S. EAGER, JR., L. JACHIMOWICZ, I. KOLODNY, and D. E. ROBINSON are with the General Cable Corporation, Bayonne, N. J.

The authors acknowledge the help and advice of A. G. Chapman, Consulting Engineer, throughout this program and the assistance of J. A. Olszewski, General Cable Corporation, in making measurements and evaluation of the data.

Vol. 7
Pt. 1
1959

on a 100-pair cable, 19 Awg copper conductors, the core being made of four units each consisting of 25 pairs having different twist lengths. The average mutual capacitance was 0.083 μf per mile and this cable henceforth will be referred to as 19 Awg 0.083 cable. The Alpeth sheath consists of a folded, corrugated aluminum tape 8 mils thick with an over-all black, high-molecular-weight polyethylene jacket. Surrounding the cable core and directly under the corrugated aluminum tape is a bedding consisting of a laminate GR-S rubber and Mylar tape having a thickness of 16 mils. Certain shielding efficiency measurements were conducted on this cable as well as on a similar cable having 2 helically applied aluminum tapes each 5 mils thick in lieu of the 8-mil folded corrugated tape.

Additional parameter tests were conducted on two similar 100-pair cables, one in size 19 Awg having an average mutual capacitance of 0.066 μf per mile and made up of six units each consisting of 16 or 17 pairs; the other in size 22 Awg having an average mutual capacitance of 0.083 μf per mile and made up of four units each consisting of 25 pairs. These will henceforth be referred to as 19 Awg 0.066 cable and 22 Awg 0.083 cable, respectively.

All of these cables were chosen as typical of normal production and were placed on wooden reels to eliminate the presence of steel in the vicinity during measurements.

Test Conditions

Most measurements of parameters were made on a reel of cable 180 feet long. Certain measurements were made on other lengths of the same cable because of measuring difficulties encountered on this length; to measure the conductance, a 20-foot length for the highest frequencies and a 3,000-foot length the lowest frequencies were employed. For most measurements it is convenient to choose a sample about 10 feet in length. However, this has the undesirable features that such a sample is insufficient in length to give representative results and that the end effects introduce errors which may make the measurements questionable. Consequently, it is necessary to choose a longer sample. However, caution must be exercised in the choice of test length with regard to wavelength in order to avoid having the open-circuit admittance appear as a short-circuit impedance and vice versa. Although such a condition can be tolerated and correct parameters determined by the methods described later, it is quite inconvenient. For these reasons a 180-foot length of cable was chosen for the bulk of the measurements. Such a length was found convenient for all frequencies studied and had the additional advantage that such a reel length could be conveniently placed in the tempeature-controlled rooms utilized in this study. The temperature range studied was from 0 F (degrees Fahrenheit) to 140 F and was maintained constant by automatic temperature controls. It is quite difficult to maintain the cable temperature precisely stable throughout its length, and for this reason the d-c resistance of the cable pair under test was periodically measured with a d-c Kelvin bridge and these measurements used to establish the average cable temperature at any time during the testing period.

Crosstalk measurements were made on a reel of cable 3,000 feet in length in order to obtain more accurate data on a long length of cable although a length shorter than 1/4 wavelength would have facilitated measurements since reflections at terminations of circuits would be only minor in nature. Each of the four units of the test length was paired and stranded in a 6,000-foot length, which is customary manufacturing procedure. Two cables were then obtained. Each was 3,000 feet in length and was composed of four units, each produced in separate runs. The test cable was kept in a shielded, insulated chamber at a controlled temperature of 68–70 F. For tests, all cable pairs were terminated in suitable networks with midpoints grounded.

Shielding studies were made at room temperature (75–85 F) on a cable length of approximately 10 feet.

Primary and Secondary Parameters

Rather than make measurements on all pairs in the cable at all temperatures and frequencies a preliminary study was made at 80 F of all pairs within the cable to establish the variation of parameters between pairs and to enable the choice of several representative pairs upon which to make the bulk of the measurements. The maximum variation at 1 kc from average values for all 100 pairs of the 19 Awg 0.083 cable for the a-c resistance was ± 1%; for mutual capacitance, ±4% in two units and ± 3% in the other two units; and for inductance, ±2%. The variation at 1 kc in conductance was not determined on the 180-foot reel because the values were less than the reliable range of the bridge. On a 3,000-foot reel of similar cable the conductance of all pairs was found to vary between 0.5 and 2.0 μmho (miromho) per mile at 1 kc. The variation between pairs of unit 3 at various frequencies was measured and the maximum variation from average values for resistance was ±1% for all frequencies; for mutual capacitance, ±4% for all frequencies; for inductance, ±2% at 1 kc and ±3% at 20 kc and above. For conductance at higher frequencies the variation was large, as it was at 1 kc.

The primary parameters (R, L, C, and G) and the secondary parameters (Z, α, β, and velocity) were obtained by measurements of the short-circuit impedance and the open-circuit admittance and by using equations 24 to 28 as indicated in Appendix II.

Resistance

The results obtained on the 19 Awg 0.083 and the 22 Awg 0.083 cables are given in Fig. 1, where the change of resistance with frequency and with temperature is shown. The resistance of the 19 Awg 0.066 cable is not plotted but the resistance of this latter cable was measured at 80 F and found to be equal to the 19 Awg 0.083 cable at low frequencies and, as would be anticipated, slightly less at the higher frequencies. The measured resistance and that computed from cable dimensions for the 19 Awg 0.066 cable at 1,000 kc were 432 and 396 ohms per mile, respectively. The change with frequency is minor up to 20 kc. In the frequency range of 20 kc to 300 kc the change is quite rapid and above 300 kc the change is very rapid. The effect of temperature on the resistance becomes less on a percentage basis as the frequency increases, as may be seen from Fig. 2 also.

The resistance increases as the frequency increases because of (1) skin effect in the individual conductors; (2) proximity effect in one conductor because of the presence of the other conductor in a

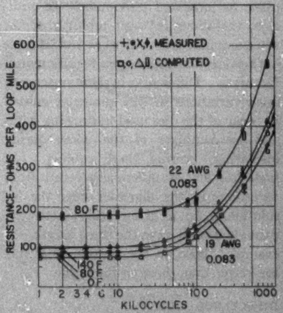


Fig. 1. Resistance per loop mile versus frequency

given pair; and (3) eddy current losses in the surrounding pairs and sheath. The a-c resistance may be computed from cable dimensions with good accuracy from the following:

$$R_{ac}=R_{dc}+\Delta R_s+\Delta R_p+\Delta R_1 \qquad (1)$$

where

R_{dc} = d-c resistance
ΔR_s = increment due to skin effect
ΔR_p = increment due to proximity effect
ΔR_1 = increment due to eddy currents in other pairs and the sheath

The component ΔR_1 is small at low frequencies and may be neglected for most practicable purposes for frequencies up to 50 kc; Appendix III gives the method by which equation 1 may be employed to estimate the a-c resistance of a pair for a cable having a folded corrugated aluminum sheath or equivalent. The estimated a-c resistances as computed from cable dimensions using equation 1 are shown in Fig. 1.

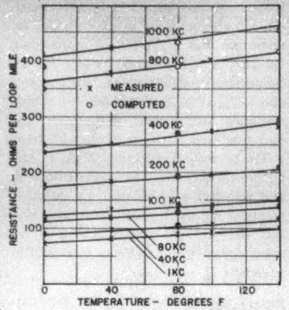


Fig. 2. Resistance per loop mile versus temperature for 19 Awg 0.083 cable

It is reasonable to expect variations in the parameters between pairs in a given cable as indicated previously. In addition, variations between cables from the same manufacturer will occur and likewise between similar cables from various manufacturers. Differences in resistance result from variations in wire diameter caused by wear of the dies used in drawing wires and by handling of the wires during insulating, pairing, and cabling. (At high frequencies differences result from variations in the diameter over the insulation.) In a given reel length the pairs do not all have the same length, for example, those on the outside of a unit are longer than those on the inside.

INDUCTANCE

The results obtained on the 19 Awg 0.083, 19 Awg 0.066, and 22 Awg 0.083 cables are given in Fig. 3 where the change of inductances with frequency is shown. Actually, the inductance changes little with frequency as compared to the changes observed in the resistance, and even less with temperature. The change with frequency is minor up to about 40 kc. In the range from 40 kc up to 1,000 kc the change is more important although at the higher frequencies the inductance is changing less rapidly because, at some frequency beyond 1,000 kc, the inductance will approach a constant value and not change with either frequency or temperature.

The inductance of the pair is reduced as the frequency increases because of skin effect in the individual wires. In addition to this, the proximity effect between the two wires of a given pair further reduces the inductance as the frequency increases. The inductance is further reduced by eddy currents flowing in adjacent pairs and in the sheath; however, both latter effects are small and only significant at frequencies above 750 kc.

The inductance of a pair of wires may be calculated[3] from the following formula:

$$L=[1.485\log_{10}2\,s/d+0.161l(x)]\times10^{-3} \qquad (2)$$

henrys per loop mile of the wire, where s is the center to center spacing of the wires of a given pair, d is the diameter of each wire, both in inches, and $l(x)$ is obtained from Fig. 24 for the applicable value of x given in equation 29. Equation 2 includes skin effect of the individual wires only. Slightly more accuracy may be obtained at higher frequencies ($x>2$) by using equation 2(A) which includes the proximity effect of the other wire of the pair.

$$L=\left[1.485\log_{10}\frac{d}{s-\sqrt{s^2-d^2}}+0.161Kl(x)\right]\times10^{-3} \qquad (2A)$$

where K is a correction factor of inductance for proximity effect and may be estimated as explained in Appendix IV. Equation 2(A) is a modified form of equation 2 and has been found to agree fairly well with the measured inductance. This equation does not include the effects of eddy currents in the surrounding wires and sheath; they reduce the inductance slightly at high frequencies. The value of inductance as given in equations 2 and 2(A) may be increased up to 2% to account for the additional length due to cabling and pairing.

The values of inductance computed from cable dimensions are given at 80 F for the three cables measured. The change of measured inductance with temperature is very small and the computed values are not shown for 0 F and 140 F although these may be estimated by using the proper values of x. Actually the only reason the inductance changes with temperature is because of an indirect effect from the change in resistance. As the resistance changes, the value of x changes and, consequently, the effects of skin and proximity effects change.

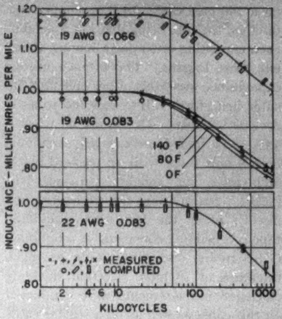


Fig. 3. Inductance per mile versus frequency

It is reasonable to expect variations in inductance due to manufacturing variations similar to those influencing the resistance. Variations in wire diameters, diameters in insulated conductors, and variations due to pairing and cabling are most important.

MUTUAL CAPACITANCE

The mutual capacitance did not vary to any significant degree over the complete temperature and frequency range studied. The maximum values on the pair tested were 0.082, 0.066, and 0.082 μf per mile for the 19 Awg 0.083, 19 Awg 0.066, and 22 Awg 0.083 cables, respectively; the corresponding minimum values were 0.081, 0.065, and 0.081. Such variations with frequency in mutual capacitance are believed to be due principally to experimental error and to a slight effect of the *GR-S* rubber-Mylar core tape. The dielectric constant of this core tape, which is placed in the cable for mechanical purposes, decreases somewhat as the frequency is increased and, since it is to a small extent in the electric field of the pair under measurement, the mutual capacitance decreases slightly as the frequency is increased.

The deviation of mutual capacitance of individual pairs from the average is due principally to their relative location with respect to the sheath, the tightness of the cabling, and variations in length of pair lays and cabling. Although these cables are fully color coded, the colors are chosen so as to not appreciably vary the dielectric constant of the polyethylene and, consequently, the mutual capacitance is not affected by this factor. The dielectric constant of polyethylene does not change significantly with temperature and with frequency over the range studied and, therefore, one would not expect the mutual capacitance to change.

The mutual capacitance of two wires in a medium may be expressed by the following well-known formula:

$$C=\frac{0.019415\times 10^{-6}\epsilon}{\log_{10} 2\, s/d}\ \text{f per mile} \tag{3}$$

where

ϵ = dielectric constant of the medium
s = center to center conductor spacing
d = wire diameter in inches

However, if the wires are surrounded by a dielectric which is the combination of polyethylene and air, this formula is not accurate unless modifications are made as indicated by Maupin[4] to calculate the effective dielectric constant of the medium surrounding the two wires. Moreover, in a cable, other wires are very near and the sheath is in the vicinity and effectively grounded, all of which results in a considerable modification of the mutual capacitance.

The following simplified formula, however, has been found useful for estimating the mutual capacitance of polyethylene insulated cables. The accuracy is generally within ±5%, which is only a little higher than the variation of capacitance between pairs found in the cable under study.

$$C=\frac{0.035}{\log_{10} 1.5\, s/d}\ \mu\text{f per mile} \tag{3A}$$

where s and d are the same as in equation 3. This formula gives a mutual capacitance of 0.085 μf per mile for the 19 Awg 0.083 cable, 0.066 for the 19 Awg 0.066 cable, and 0.082 for the 22 Awg 0.083 cable studied.

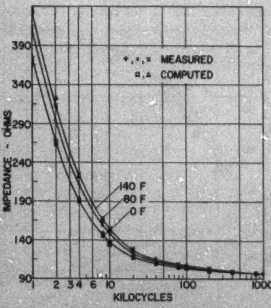


Fig. 4. Characteristic impedance, magnitude versus frequency for 19 Awg 0.083 cable

CONDUCTANCE

In polyethylene-insulated telephone cables, the conductance (also called mutual conductance) is extremely small and, in fact, may be neglected entirely when calculating the secondary parameters for all except frequencies in excess of 750 kc. For a normal cable, therefore, G is so small that it has no effect on the secondary parameters: impedance, velocity, phase shift, and attenuation. It was found difficult to obtain consistently reliable values of G when using equation 21 of Appendix II. Consequently, at the sacrifice of some accuracy, the conductance was measured on suitable lengths so that the open-circuit impedance yielded the conductance directly. Since the variation of conductance between pairs was large, care was taken to measure sufficient pairs to assure the tested pair was near the average. (The results obtained on the 19 Awg 0.083 cable are given in Fig. 10, where the change of conductance with frequency is shown along with values computed from cable dimensions using equation 4(A).) The change with temperature is very minor and has been neglected. The measured conductance of the 19 Awg 0.066 cable varied from about 1 μmho per mile at 1 kc to 290 μmho per mile at 1,000 kc and that of the 22 Awg 0.083 cable varied from about 1 μmho per mile at 1 kc to 330 μmho per mile at 1,000 kc. These values are valid throughout the temperature range of 0 F to 140 F.

The conductance of two wires in a medium may be expressed in accordance with the following formula:

$$G=6283\, fC\,(pf)\ \mu\text{mho per mile} \tag{4}$$

where

pf = power factor of medium
C = capacitance, μf per mile
f = frequency, kc

If the medium is a combination of polyethylene and air and other insulating materials, such as *GR-S* rubber-Mylar tape, are present, and if some contamination is present, it is impossible to derive an exact formula for calculating conductance. The power factor (pf) of colored polyethylene used in this cable varies from 0.0002 to 0.0005, depending on color. Experience has indicated, however, that the effective pf between wires in polyethylene-insulated paired cables is higher because of the afore-mentioned tape and contaminating particles. Consequently, if the capacitance (C) of equation 4 is replaced with that of equation 3(A) the following formula is very useful in estimating G.

$$G=\frac{0.154f}{\log_{10} 1.5\, s/d}\ \mu\text{mho per mile} \tag{4A}$$

where f is the frequency in kc, and s and d are the same as in equation 3. In this formula it is assumed that the conductance is not affected by temperature over the range of 0 F to 140 F. However, the variation of conductance between pairs is large as indicated previously and this equation must be used with this in mind.

IMPEDANCE

The characteristic impedance may be expressed as

$$\sqrt{\frac{R+j\omega L}{G+j\omega C}}$$

as given in equation 19 of Appendix II. For polyethylene-insulated telephone cable, G is negligible compared to $j\omega C$. At low frequencies, $j\omega L$ is negligible compared with R and the impedance is $\sqrt{R/j\omega C}$. At high frequencies, R is negligible compared with $j\omega L$ and the impedance is $\sqrt{L/C}$.

The magnitude of the impedance obtained on the 19 Awg 0.083 cable is given in Fig. 4 along with the values computed from cable dimensions at 0 F and 140 F obtained by utilizing primary parameters computed from cable dimensions and equation 19. The real and imaginary measured components are given in Fig. 5. The results obtained on the 19 Awg 0.066 and the 22 Awg 0.083 cables are given in Table I. The impedance decreases rapidly as the frequency is increased until

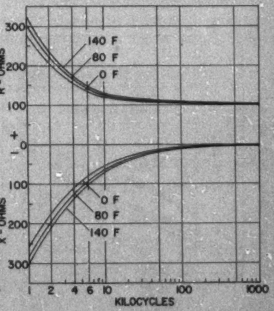


Fig. 5. Characteristic impedance, $Z_0=R-jX$ versus frequency for 19 Awg 0.083 cable

Table I. Measured and Computed Secondary Parameters at 80 F

Frqeuency, Kc	Characteristic Impedance, Z, Ohms (Angle in Degrees)		Attenuation, α Db per Mile		Phase Change, β Radians per Mile		Velocity, VP, Miles per Second × 10³	
	Measured	Computed	Measured	Computed	Measured	Computed	Measured	Computed
19 Awg 0.066 μf per Mile Cable								
1	464/−42.6	470/−42.6	1.12	1.11	0.141	0.139	44.5	45.2
2	330/−40.2	334/−40.2	1.53	1.51	0.208	0.205	60.3	61.2
4	239/−35.6	241/−35.7	1.99	1.97	0.319	0.315	78.6	79.9
8	181/−28.0	182/−28.2	2.43	2.42	0.526	0.517	95.5	97.2
10	169/−25.1	169/−25.2	2.56	2.53	0.629	0.618	99.8	101.6
20	147/−15.8	147/−15.9	2.86	2.82	1.16	1.14	108.4	110.7
40	139/−9.36	138/−9.25	3.23	3.12	2.25	2.19	111.6	114.7
80	135/−6.04	135/−5.95	4.04	3.92	4.39	4.30	114.5	116.8
100	134/−5.33	134/−5.31	4.44	4.35	5.46	5.36	115.1	117.2
200	130/−3.84	130/−3.90	6.20	6.26	10.6	10.5	117.8	119.9
400	128/−2.65	128/−2.83	9.23	8.96	21.0	20.6	119.6	122.0
800	125/−2.17	126/−1.99	12.7	12.4	41.1	40.7	122.3	123.5
1,000	123/−1.96	126/−1.77	15.5	13.9	50.8	50.8	123.5	123.6
22 Awg 0.083 μf per Mile Cable								
1	590/−44.0	594/−44.0	1.83	1.83	0.219	0.217	28.7	28.9
2	418/−42.9	420/−43.0	2.54	2.53	0.314	0.312	39.9	40.2
4	297/−40.9	298/−41.0	3.48	3.47	0.461	0.458	54.4	54.9
8	214/−37.0	214/−37.1	4.59	4.58	0.700	0.695	71.8	72.4
10	194/−35.2	194/−35.3	4.98	4.96	0.810	0.803	77.6	78.2
20	148/−27.5	147/−27.6	6.09	6.05	1.34	1.33	93.6	94.5
40	125/−18.2	124/−18.2	6.96	6.86	2.42	2.39	103.6	105.1
80	116/−11.4	116/−11.4	8.13	8.06	4.62	4.59	108.8	109.1
100	114/−9.79	113/−9.43	8.66	8.56	5.74	5.68	109.5	110.6
200	110/−6.65	109/−6.56	11.3	11.1	11.1	11.1	112.9	113.7
400	106/−4.88	106/−4.90	16.5	16.2	21.6	21.6	116.4	116.6
800	101/−3.66	103/−3.63	23.8	23.5	41.8	40.9	120.2	119.5
1,000	101/−3.43	103/−3.29	27.3	26.4	51.7	52.4	121.5	119.9

about 100 kc after which the decrease is minor. The impedance decreases rapidly because the component $j\omega C$ increases directly with frequency whereas the resistance changes very little, and, at low frequencies $j\omega L$ is small compared with the resistance. At the higher frequencies, the resistance is small compared with $j\omega L$, hence the decrease in impedance is due only to the slight decrease in the inductance. Since L, C, and G are affected little by temperature and R is appreciably affected by temperature, the impedance varies considerably from 0 F to 140 F at the low frequencies, whereas above approximately 20 kc the temperature has very little effect on the impedance; indeed, above 100 kc this effect may be considered negligible.

ATTENUATION

The results obtained on the 19 Awg 0.083 and 22 Awg 0.083 cables are given in Fig. 6, where the change of attenuation with frequency is shown. The results obtained on the 19 Awg 0.066 cable are given in Table I. The change of resistance of the conductors with frequency is the major contributor to the change of attenuation. The conductance of normal polyethylene-insulated cables is so low that it does not contribute to attenutation to any significant degree except at frequencies above 750 kc. The effect of temperature on attenuation is equally noticeable throughout the frequency range. Although the resistance is affected more by temperature at the lower frequencies than at the higher ones, the attenuation varies as the square root of the resistance at low frequencies but the loss contributed by series resistance varies directly with resistance at high frequencies. This results in the temperature having a more uniform effect on attenuation.

Once the primary parameters R, L, C, and G have been computed from cable dimensions, the attenuation may be computed by using equation 22 of Appendix II. In Fig. 6 and Table I, the values of attenuation computed from cable dimensions given are from equation 22.

PHASE CHANGE: VELOCITY

The results obtained on the 19 Awg 0.083 cable are given in Fig. 7 and on the 19 Awg 0.066 and 22 Awg 0.083 cables in Table I. The phase change increases rapidly with frequency; it is affected slightly by temperature.

By utilizing equation 23 of Appendix II the phase change may be computed from cable dimensions from the primary parameters previously computed. These computed values are shown for 80 F in Fig. 7 and Table I.

The velocity of propagation (VP) in miles per second is related to the phase change by the following relationship:

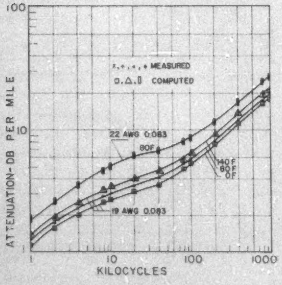


Fig. 6. Attenuation in db per mile versus frequency

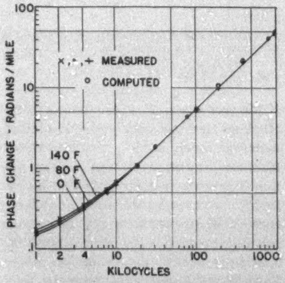


Fig. 7. Phase change versus frequency for 19 Awg 0.083 cable

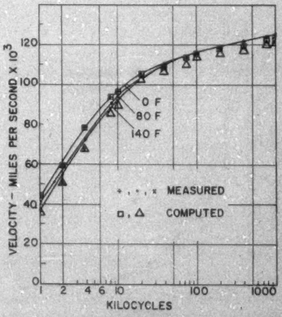


Fig. 8. Velocity versus frequency for 19 Awg 0.083 cable

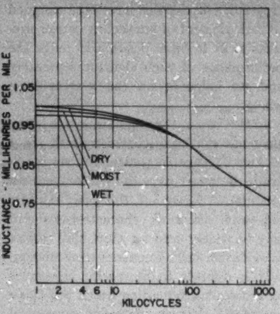


Fig. 9. Inductance per mile versus frequency for dry, moist, and wet 19 Awg 0.083 cable

$$VP = \omega/\beta = 6.283\,f/\beta \qquad (5)$$

where

f = frequency, kc
β = phase change, radians per mile

The phase change and the velocity do not vary significantly with temperature. The measured values of the velocity for the 19 Awg 0.083 cable are given in Fig. 8 along with the velocities computed from cable dimensions at 0 F and 140 F, and the measured values for the 19 Awg 0.066 and 22 Awg 0.083 cables are given in Table I along with the values computed from cable dimensions.

EFFECTS OF MOISTURE ON CABLE PARAMETERS

Moisture, upon entering a polyethylene-insulated telephone cable, may not render it inoperative, but does modify its transmission parameters. A study was made of the transmission parameters of the 100-pair 19 Awg 0.083 cable completely filled with water. The cable parameters were first measured on the cable as received from the factory. The cable was then completely flushed for 24 hours with tap water, having a conductivity of about 70 μmho-centimeters at 1 kc and 77 F, and allowed to remain filled with the flushing water. After the parameters were determined with the cable filled with water, the water was allowed to drains out by uspending the cable in a vertical position for several hours. Immediately following this, the parameters were measured again while the cable was still moist. These latter measurements were completed within 6 hours after cable was drained. The cable was then dried again for several days by pumping dry nitrogen through it. The parameters were again measured and found to be exactly the same as those previously measured when the cable was in the "as received" condition. All measurements were made at room temperature which was within the range of 75 to 85 F. The results are given in Figs. 9–14.

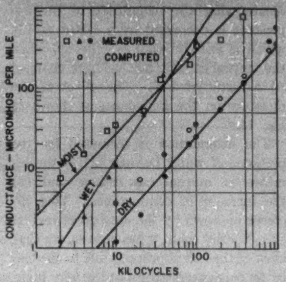


Fig. 10. Conductance per mile versus frequency for dry, moist, and wet 19 Awg 0.083 cable

EFFECTS OF MOISTURE ON PRIMARY PARAMETERS

The results indicate that the water has virtually no effect on the series resistance, as would be anticipated, and no effect on the series inductance except perhaps in the 1–5-kc range as may be seen in Fig. 9 where the inductance is slightly lower wet than dry. This is because at these low frequencies the water has sufficient conductivity so that the cable pair acts somewhat as if it were a twin coaxial. At higher frequencies it does not act in this manner.

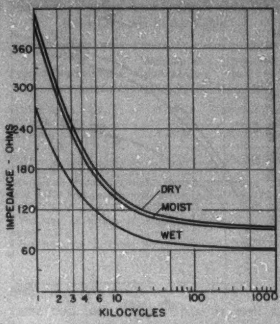


Fig. 11. Characteristic impedance, magnitude versus frequency for dry, moist, and wet 19 Awg 0.083 cable

The mutual capacitance of the wet cable was found to be 0.186 μf per mile which is 2.25 times as high as that of the dry cable which indicates that the dry mutual capacitance is approximately 25% that of the self-capacitance of one insulated conductor. This is to be anticipated because the computed self-capacitance of one insulated conductor is 0.374 μf per mile, and the wet mutual capacitance is equal to one half the self-capacitance (capacitance to water) of a single insulated conductor. This is not the case, however, if the cable has become only moist because in this case the mutual capacitance was found to be 0.090 μf per mile as compared with the dry value of 0.082. The wet and moist mutual capacitance did not change with frequency.

The conductance is increased by the presence of water as indicated in Fig. 10. Reference to equation 4 would indi-

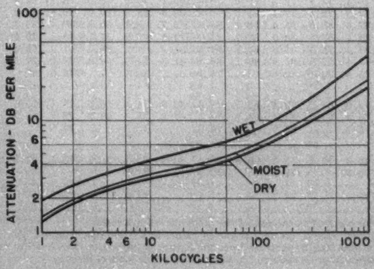


Fig. 12 (left). Attenuation in db per mile versus frequency for dry, moist, and wet 19 Awg 0.083 cable

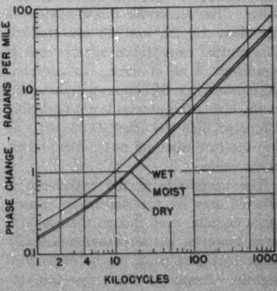


Fig. 13 (right). Phase change versus frequency for dry, moist, and wet 19 Awg 0.083 cable

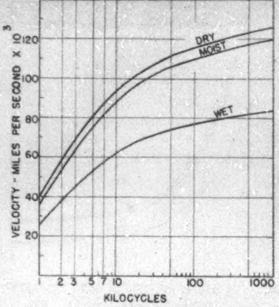


Fig. 14. Velocity versus frequency for dry, moist, and wet 19 Awg 0.083 cable

cate that the conductance would increase by a factor of 2.25 because of the mutual capacitance only. Actually, the wet conductance increases more than this factor especially at frequencies above 10 kc. This is because the water does not have sufficient conductivity to act as a metallic conductor except to some extent at the very lowest frequencies, hence there is a large dielectric loss thus increasing the pf. As indicated previously at low frequencies, the water has sufficient conductivity so that the inductance is decreased. The increase in conductance is accentuated at the highest frequencies where the water behaves as a dielectric material having a high dielectric loss. Since the conductance has so little effect on the transmission properties of the cable, this increase is not important except at very high frequencies.

EFFECTS OF MOISTURE ON SECONDARY PARAMETERS

Although only the mutual capacitance and conductance are significantly changed by the presence of water, all of the secondary parameters—attenuation, impedance, phase shift, and velocity—are significantly changed by the presence of water. The reason for this is obvious if equations 19, 22, and 23 are studied. The mutual capacitance appears and is significant in all of these. In addition, the conductance that appears in these and at the higher frequencies is not negligible compared with the other components as was the case when no water was present in the cable.

The magnitude of the impedance is affected only by the increase in capacitance because, even though the conductance, however, has increased sufficiently to decrease the angle of the impedance slightly. Examination of the data indicates that the magnitude of the impedance is decreased by a factor of 1.5, which is exactly what is expected when the mutual capacitance has been changed by a factor of 2.25. The effect of the moisture on the impedance is shown in Fig. 11.

The attenuation is increased by the presence of water because of (1) the increase of mutual capacitance and (2) the increase of conductance. This increase is shown by the data given in Fig. 12. For dry cable, it has been shown that most of the attenutation is due to the loss in the conductors except at the very highest frequencies. This is not true when water is in the cable.

The phase change and the velocity are both affected slightly in so far as slight amounts of moisture in the cable are concerned; however, in the case of the wet cable, both of these parameters are significantly affected. This may be seen clearly from Figs. 13 and 14.

Crosstalk

Crosstalk and noise currents in telephone circuits are contributed by many sources. In addition to currents flowing through unbalances of cable pairs, there are secondary inductions from inside plant at exchange offices, crosstalk in repeaters, voltages induced by open wires, and so on. Accordingly, filters, balancing, compandors, and other means are used to suppress the effect of crosstalk and noise currents. Target values of minimum crosstalk coupling loss have been assigned to each element of paralleling carrier systems so as to be in position to judge the performance of each element separately and its influence on the system as a whole.

In this paper, the crosstalk in one of the elements will be discussed, namely, in a 19 Awg 0.083 cable. Although not specifically designed as a toll cable from an attenuation viewpoint, this cable exhibits toll grade crosstalk characteristics and may be widely used for short-haul carrier as well as for voice-frequency transmission.

It is difficult to set specific objectives for crosstalk attenuation which this cable must fulfill. It is a matter of economical balance between cable cost and the costs of carrier systems including means employed to improve the over-all performance of the system. Therefore, the crosstalk characteristics of a modern polyethylene-insulated cable will be presented and for reference only some of the figures of crosstalk attenuation found on cables presently in use on carrier services will be quoted.

Some published data show[8] that in paper cables the rms value of far-end crosstalk coupling loss in a mile of quadded cable consisting of 37 quads (excluding strong couplings between pairs within a quad) has been found to be 75 db (decibels) at 165 kc. This high level of crosstalk loss permitted the use of quadded cables in lengths up to 200 miles without pair balancing, for *N1* type carrier designed to prevent near-end crosstalk

Table II. Far-End Output to Output Crosstalk Coupling Loss at 150 Kc in 3,000 Feet 19 Awg 0.083 Cable for Crosstalk* Between Pairs in the Same Unit

All Four Units Identical, No Like Pair Twists in a Unit, Each Unit Consists of Three Layers, 3, 8, and 14 Pairs

Type of Combination	Number of Combinations in Each Unit	Crosstalk Coupling Loss, Db									
		Unit 1		Unit 2		Unit 3		Unit 4		In All Units	
		Rms	Min.	Rms	Min.	Rms	Min.	Rms	Min.	Rms	Min.
Within layers:											
Adjacent pairs	25	69.6	57.0	75.0	70.0	71.1	60.0	70.2	59.2	71.0	57.0
Alternate pairs	22	78.1	71.1	79.3	69.5	76.4	69.0	79.6	70.0	78.2	69.0
Two separators	22	81.3	72.1	79.9	71.0	78.8	68.4	78.3	70.0	79.4	68.4
Three separators	18	87.6	82.5	85.6	78.7	85.0	75.7	88.4	80.6	86.4	75.7
Four separators	14	88.1	80.6	89.9	83.1	89.0	79.5	89.0	81.5	88.9	79.5
Five separators	14	87.3	78.8	97.6	91.7	91.6	81.7	95.5	88.1	91.2	78.8
Six separators	7	79.4	72.5	89.8	85.1	86.2	81.3	91.6	84.9	84.0	72.5
Rms	122	75.5		79.6		76.3		76.1		76.6	
Adjacent layers:											
Center to first layer	24	72.0	61.0	73.4	62.1	72.8	60.6	74.8	66.6	73.1	60.6
First to second layer	112	76.4	63.9	81.4	68.7	78.1	67.3	78.6	66.7	78.3	63.9
Rms	136	75.2		78.5		76.5		77.6		76.8	
Alternate layers:											
Center to second layer	42	80.6	72.4	75.0	60.3	80.4	70.8	81.4	74.6	78.5	60.3
Units rms	300	75.8		78.1		76.8		77.3			
All units rms	1200									76.9	
Rms in a mile† of cable		73.4		75.7		74.4		74.9		74.5	

* The rms value is the number of db corresponding to the rms of the crosstalk voltage ratios.

† All crosstalk values, except where expressly stated, are the measured values in a 3,000-foot length. To scale these values down to crosstalk in a mile length, subtract from the figures given in the table 2.4 db. This assumes that crosstalk voltages increase in proportion to the square root of the ratio of lengths.

Table III. Far-End Crosstalk Loss as in Table II for Crosstalk Between Pairs in Different Units

Type of Combination, Layers	Number of Combinations	Crosstalk Coupling Loss, Db			
		Adjacent Units 1 to 4		Opposite Units 1 to 3	
		Rms	Min.	Rms	Min.
Second to second:					
Likes	14	76.1	69.9	86.1	79.3
Unlikes	182	83.6	76.0	93.8	85.1
Second to first:					
Unlikes	224	91.8	83.4	102.5	93.9
Second to center:					
Unlikes	84	97.0	85.8	109.9	103.6
First to first:					
Likes	8	81.2	74.2	97.4	91.1
Unlikes	56	97.4	90.2	103.5	92.1
First to center:					
Unlikes	48	100.3	89.5	107.7	102.0
Center to center:					
Likes	3	87.3	82.8	91.0	87.0
Unlikes	6	111.9	106.5	114.0	111.7
Total, Rms	625	86.6		96.9	
Rms* in a mile of cable		84.2		94.5	

* See footnote 2, Table II.

Table IV. Far-End Loss Crosstalk as in Table II for Crosstalk in the Cable

Type of Combinations	Number of Combinations	Crosstalk in CU*
Within units	1,200	24,331,700
Opposite units	1,250	255,200
Adjacent units	2,500	5,483,200
Total	4,950	30,070,100
Crosstalk rms in CU[8]	6,075	

Grand rms crosstalk coupling loss in the cable: db in a 3,000-foot length = 82.2; db in a mile† = 79.8

* For the definition of crosstalk units (CU) see Appendix V.

† See footnote 2, Table II.

from being heard. Paired cables of older (not staggered pairs) design were not as good. The rms crosstalk loss was found to be as low as 55 db in a mile. Improved unit-type cables have had much better rms crosstalk loss, namely, 68 db in a mile of 25-pair unit, with rms coupling loss between units about 9 db higher. These are the crosstalk values which will be compared with the values found in the cable under study. Since some carrier systems permit near-end crosstalk to be heard, values of near-end as well as far-end crosstalk coupling loss are presented. Far-end crosstalk was measured as output to output or equal level crosstalk coupling loss.

Measurements of crosstalk can be very much affected by reflections from mismatched terminations, and a special study was conducted to determine whether pairs not under direct test should be left floating or terminated. It was concluded that for suppression of reflections from open ends of pairs, phantoms, and other tertiary circuits, it is best to terminate all pairs and ground their center taps.

Crosstalk energy is transferred from the disturbing to the disturbed circuits by electrostatic or electromagnetic couplings, which are the results of one or another sort of assymetry in the geometrical shape of pairs introduced by the necessity of design and by unavoidable irregularities in manufacture of cable. Some of the theoretical considerations of the mechanism of this transfer of energy by transverse and interaction paths, as well as formulas used in computations, are given in Appendix V.

Measured Crosstalk Coupling Loss

At 150 kc, representative for medium-frequency range, far-end crosstalk was measured on all units of the cable. The results are given in Tables II-IV. Tests were conducted on the basis that similar results are obtained for a given set of two pairs if one is made disturbing and the other disturbed, and vice versa. A limited number of pair combinations was tested to verify this assumption. Strictly speaking, such reciprocity does not exist for far-end crosstalk[8] due to the difference in the tertiary path of the two pairs in question. The magnitude of difference involved was found to be small for the controlling crosstalk combinations and, therefore, only one test per combination of two pairs was performed.

In Tables II–IV the rms values of cross-

Table V. Far-End Output to Output Crosstalk Coupling Loss 1–1,000 Kc for Pairs Within Layers

Unit 3 in 100-Pair No. 19 0.083 Cable, Db in 3,000-Foot Length

Kc	Number of Separators													
	Adjacent		Alternate		2		3		4		5		6	
	Number of Combinations													
	25		22		22		18		14		14		7	
	Rms	Min.	Rms	Min.	Rms	Min.	Rms	Min.	Rms	Min.	Rms	Min.	Rms	Min.
1	100.8	85.4	107.4	99.7	111.9	102.3	125.0	117.5	136.8	131.5	144.9	133.0	133.0	126.2
2	98.1	91.8	104.3	97.4	109.2	99.4	121.1	114.3	130.1	122.1	135.0	127.0	125.2	121.2
4	95.3	89.1	101.6	94.7	106.0	96.0	115.8	109.0	123.2	116.5	125.3	119.3	117.6	113.2
8	92.3	84.5	98.2	92.8	102.2	92.6	109.9	101.8	115.7	106.6	116.7	107.8	110.3	104.1
15	88.3	78.4	94.3	87.8	97.4	94.1	103.8	94.8	108.7	99.2	110.0	100.5	104.3	99.6
30	83.5	72.9	89.1	81.5	92.1	82.1	97.8	88.4	102.3	91.6	103.6	93.9	97.8	93.2
60	78.3	67.4	83.7	75.8	86.5	76.3	92.4	83.1	96.1	86.8	98.3	88.5	93.0	88.4
100	74.4	63.2	79.6	71.7	82.1	72.1	88.5	79.0	92.2	82.7	94.0	84.7	89.3	84.4
150	71.1	60.0	76.4	69.0	78.8	68.4	85.0	75.7	88.4	79.5	91.6	81.7	86.2	81.3
200	69.2	58.2	73.0	65.3	76.8	67.0	83.2	74.5	86.9	78.2	88.0	79.7	84.5	80.0
250	66.8	55.5	71.9	64.2	74.8	64.6	81.1	72.9	84.4	75.9	86.4	77.9	82.2	77.7
300	64.7	53.5	70.0	62.6	72.2	62.9	79.2	70.9	83.6	73.1	83.5	75.7	80.3	76.2
400	61.0	49.6	67.1	59.4	69.4	59.2	76.0	68.2	80.5	69.0	81.6	72.6	77.0	72.8
500	60.1	49.0	65.6	58.3	68.3	58.1	73.9	66.1	78.6	69.2	80.0	71.7	75.8	71.8
600	58.9	47.5	64.2	56.9	67.0	56.9	73.5	64.4	77.3	68.8	77.9	70.1	75.3	71.3
700	57.3	45.8	62.6	55.3	65.4	55.5	71.8	63.1	75.6	66.4	77.7	69.6	73.2	69.9
800	55.3	43.8	60.8	53.8	63.8	53.6	70.3	61.6	72.8	64.3	75.6	67.9	71.7	67.7
900	54.9	43.7	60.3	53.4	63.1	53.3	70.3	60.8	73.1	62.8	74.0	67.2	71.1	66.9
1,000	54.2	43.3	60.0	52.2	62.2	51.7	68.7	60.0	70.3	63.0	73.3	66.0	69.8	66.4

talk, that is, the average crosstalk power ratios, are given for every group of pair combinations in a unit, between units and for the cable as a whole. This tabulation purposely includes the values measured in the 3,000-foot length. Values computed for a mile length are given separately. It was intended to treat the summation of crosstalk currents with length of cable separately from the data derived by the direct test.

Computed for a mile length of cable, the rms far-end crosstalk coupling loss at 150 kc was 74.5 db for pair combinations within a unit, 84.3 db and 94.6 db for combinations between adjacent and opposite units, respectively, and 79.8 db for all combinations within the cable.

Far-end and near-end crosstalk coupling loss as a function of frequency in the range from 1 to 1,000 kc was measured in unit 3. This unit was selected as having rms value of crosstalk loss closest to the grand rms of all four units (Table II).

Table VI. Far-End Crosstalk Coupling Loss as in Table V for Pairs in Different Layers

	Layers						Grand Rms Whole Unit	
	Adjacent				Alternate, Center to Second			
	Center to First		First to Second					
	Number of Combinations							
	24		112		42			
Kc	Rms	Min.	Rms	Min.	Rms	Min.	Number of Combinations	Rms
1	104.0	93.2	108.3	98.5	111.5	102.4	300	107.2
2	101.3	90.6	105.9	96.3	109.0	99.7	300	104.7
4	98.3	87.8	102.9	93.3	106.1	97.3	300	101.8
8	91.5	82.9	100.1	91.3	102.7	93.4	300	98.7
15	90.1	78.3	96.6	86.5	98.5	90.7	300	93.8
30	85.1	73.1	91.5	80.1	93.3	84.7	300	90.0
60	79.9	67.7	86.0	76.9	88.0	78.7	300	84.8
100	75.9	63.6	82.2	71.0	83.9	74.4	300	80.8
150	72.8	60.6	78.1	67.3	80.4	70.8	300	75.8
200	71.0	58.2	75.9	64.9	78.1	68.8	300	73.6
250	68.9	56.5	74.3	63.6	76.1	66.7	300	71.8
300	66.7	54.9	72.0	61.1	74.6	65.4	300	69.6
400	63.2	52.5	69.1	58.1	72.2	62.8	300	66.4
500	62.3	50.3	67.7	57.0	70.1	60.5	300	65.2
600	60.8	48.6	66.2	55.6	68.7	58.9	300	63.8
700	59.5	47.2	64.4	53.9	67.0	58.1	300	62.2
800	57.7	45.5	62.4	52.0	65.0	55.5	300	60.2
900	57.4	45.1	61.7	51.5	64.2	55.6	300	59.5
1,000	56.8	44.3	61.3	50.6	63.9	55.0	300	59.1
							5,700	

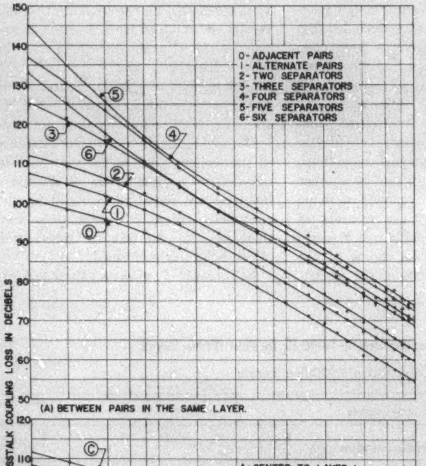


Fig. 15. Far-end output to output rms crosstalk coupling loss versus frequency in unit 3 of 3,000-foot length of 19 Awg 0.083 cable

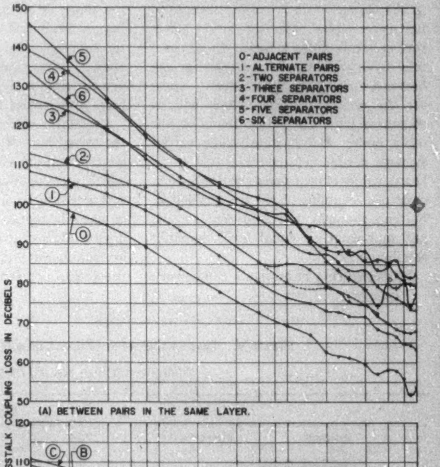


Fig. 16. Near-end rms crosstalk coupling loss versus frequency in unit 3 of 3,000-foot length of 19 Awg 0.083 cable

Vol.
Pt. 1
195

Table VII. Near-End Crosstalk Coupling Loss From 1–1,000 Kc for Pairs Within Layers

Unit 3 in 100-Pair No. 19 0.083 Cable, Db per 3,000-Foot Length

	Number of Separators													
	Adjacent		Alternate		2		3		4		5		6	
	Number of Combinations													
	25		22		22		18		14		14		7	
Kc	Rms	Min.	Rms	Min.	Rms	Min.	Rms	Min.	Rms	Min.	Rms	Min.	Rms	Min.
1	101.3	95.1	108.5	100.2	112.9	102.1	126.6	118.8	138.9	131.5	145.8	139.2	133.7	125.4
2	98.6	92.4	106.0	97.9	110.4	99.7	123.8	116.4	133.7	126.4	136.1	129.6	125.7	117.5
4	94.8	88.8	102.9	94.4	107.4	97.3	118.9	113.3	126.4	117.5	126.9	121.6	119.1	111.6
8	89.2	83.3	98.6	89.8	104.3	94.9	111.7	103.6	117.0	108.5	117.9	108.6	112.6	107.1
15	83.7	77.5	93.5	84.5	99.0	90.6	105.6	96.9	110.8	101.4	111.2	101.1	107.0	103.7
30	77.8	72.5	87.0	80.5	92.2	83.7	100.3	92.2	105.2	95.7	104.2	93.7	101.7	98.1
60	72.6	67.0	80.2	71.0	85.6	76.5	96.6	89.5	101.6	95.4	98.3	87.6	98.7	94.9
100	69.2	64.6	76.3	68.2	85.0	77.8	90.8	82.3	98.6	90.0	95.9	85.7	97.5	91.0
150	67.1	60.2	75.0	67.2	83.6	75.1	87.8	82.6	91.7	88.5	94.7	86.5	90.7	84.9
200	62.6	56.3	73.1	67.8	79.7	69.5	87.7	83.8	88.9	87.1	93.5	88.1	85.8	78.1
250	61.7	53.7	72.8	66.0	78.1	68.2	85.7	81.2	88.0	83.6	90.9	81.3	83.4	75.8
300	61.2	53.6	71.8	63.4	75.5	64.8	83.6	76.4	88.3	85.2	87.6	78.2	81.2	73.4
400	59.5	52.5	71.7	65.9	74.6	64.9	83.0	74.0	85.8	80.5	88.2	83.0	78.7	71.1
500	57.2	50.5	68.4	62.5	71.9	64.0	79.3	73.6	86.1	82.3	83.7	75.0	74.6	66.5
600	58.2	52.1	67.4	60.1	70.1	63.9	77.5	72.4	85.8	78.6	84.3	79.3	80.0	73.2
700	57.8	52.1	66.7	61.6	68.4	59.0	76.4	70.6	82.3	75.6	85.9	81.3	79.1	74.0
800	55.8	49.5	64.7	61.5	67.7	56.8	75.2	70.2	80.9	75.9	82.0	73.0	80.1	74.8
900	51.9	45.4	64.3	57.6	67.5	59.6	73.7	65.8	80.0	73.3	81.6	74.0	74.6	73.9
1,000	53.8	46.0	63.3	57.6	68.0	60.9	73.5	66.0	79.2	73.5	82.5	76.0	77.2	70.9

In Tables V and VI, Figs. 15 and 15(A), test data are presented for far-end and in Tables VII and VIII and Figs. 16 and 16(A) for near-end crosstalk coupling loss.

Introduction of 25 different twists in a 25-pair unit eliminated the source of high crosstalk at carrier frequencies between nonadjacent pairs. It can be seen from the tables that the rms crosstalk coupling loss grows with increased separation between pairs. It is worst for adjacent pairs, followed by combination of center to first layer, and so on. This pattern is followed in all units, for their rms as well as for the worst values. This regularity can be utilized for selection, for carrier transmissions,[7] of pair combinations having desired level of crosstalk coupling loss, when not all of the pairs are used for carrier transmission. At voice frequency (1 kc) the rms crosstalk coupling loss for the worst combination, that is, between adjacent pairs, was over 100 db in a 3,000-foot reel, which is much better than values found in exchange area cables originally designed for voice frequency and in use for many years.

The curves of far-end rms crosstalk coupling loss are of regular shape. Strong currents of transverse crosstalk arrive at the far end in phase and the magnitude of resulting currents changes with frequency in a regular fashion. Weaker currents of interaction crosstalk arrive at far end with phase relation changing with frequency. In the individual pair combinations which, due to chance occurrence, have weak transverse coupling, interaction currents influenced the resultant crosstalk thus causing irregularities in shape of curves. Crosstalk currents between pairs linked by strong transverse coupling were influenced by interaction crosstalk to a small degree and curves of crosstalk loss were regular.

At near end, individual crosstalk currents do not arrive in phase and this summation is random. Since there is a multitude of unbalances of a positive and negative sign randomly distributed along the cable, the cancellation or summation of crosstalk currents, each traveling to the near end along a different path length, will depend on the phase shift between currents. As the frequency increases, the distance between unbalances and the length of cable expressed in wavelength changes, causing currents arriving at the near end to add and subtract periodically. Two cases of near-end crosstalk coupling loss between two pairs are shown in Fig. 17 for mixed couplings (adjacencies) and purely inductive coupling of pairs (six separators). Note that the curves are not wavy as long as the cable is electrically short. At 100–300 kc (curve *B*) some couplings are at such a distance from each other that subtraction of currents prevails. At a higher frequency, the distance in wavelength between individual couplings

Table VIII. Near-End Crosstalk Coupling Loss as in Table VII for Pairs in Different Layers

	Layers							
	Adjacent				Alternate, Center to Second			
	Center to First		First to Second					
	Number of Combinations						Grand Rms Whole Unit	
	24		112		42			
Kc	Rms	Min.	Rms	Min.	Rms	Min.	Number of Combinations	Rms
1	105.0	93.4	109.0	96.6	110.7	99.5	300	108.0
2	102.5	91.2	106.5	94.3	108.2	97.0	300	105.4
4	99.8	94.4	103.4	90.6	104.7	93.4	300	102.0
8	95.3	87.0	98.9	87.3	100.8	89.7	300	97.0
15	91.6	84.1	93.7	82.1	96.5	83.9	300	91.2
30	85.6	80.1	87.4	76.6	91.6	77.8	300	85.9
60	79.5	71.6	83.1	70.8	86.3	73.0	300	80.7
100	76.0	65.7	77.9	68.0	80.4	70.2	300	76.2
150	74.3	62.7	69.5	62.5	79.3	69.3	300	71.7
200	70.4	61.2	73.1	60.9	77.0	68.4	300	71.0
250	68.8	60.5	72.9	62.1	75.3	64.4	300	70.3
300	68.0	59.0	71.6	60.3	74.2	63.0	300	69.5
400	66.7	60.3	70.0	61.4	73.7	65.7	300	68.0
500	67.3	61.8	68.4	57.1	71.9	64.5	300	66.0
600	66.5	59.9	64.6	53.1	70.8	62.2	300	65.1
700	65.1	57.4	64.7	55.3	67.9	60.1	300	64.7
800	63.1	56.1	63.5	52.0	60.6	57.9	300	63.0
900	62.9	55.4	65.6	54.8	66.1	57.7	300	61.2
1,000	63.0	57.5	61.7	51.2	64.3	53.6	300	61.3
							5,700	

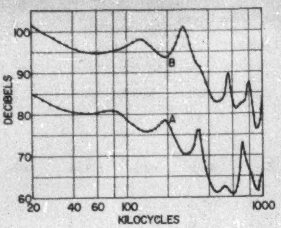


Fig. 17. Near-end crosstalk coupling loss between individual pairs versus frequency

A—Between adjacent pairs
B—Between pairs having six separators

is changed and periodical summation and subtraction of currents takes place as frequency is increased.

Curves of rms values of near-end crosstalk coupling loss, shown in Figs. 16(A) and 16(B), are wavy in a less pronounced way. If the number of pairs in a sample is increased, the rms curves will become still more regular as shown by a dotted line for pairs with two separators at frequencies of 60–500 kc, plotted with the size of the sample increased four times to include pairs in all units of the cable.

CHANGE OF CROSSTALK WITH FREQUENCY

Measured and Computed

It is of interest to analyze whether measuring crosstalk at one frequency is sufficient for a reasonably accurate prediction of its value in the whole frequency range.

The rms values of crosstalk coupling loss are dominated by the pairs contributing strongest crosstalk currents, on which interaction crosstalk has little effect. It could be expected, therefore, that changes with frequency can be computed using the classical Campbell's formula for crosstalk ratio, equation 40, Appendix V.

It follows from this equation that far-end output to output crosstalk coupling loss expressed in db is

$$F_l = 20 \log_{10}\left[\frac{j\omega C_u Z_o}{8} - \frac{j\omega L_m}{2Z_o}\right] \qquad (6)$$

where

F_l = far-end coupling loss, db
C_u = capacitance unbalance in length l, farads
L_m = mutual inductance in length l, henrys
Z_o = characteristic impedance of the line, ohms

Between far separated pairs, the capacitive coupling is almost screened out by intervening pairs, and the first term of the equation vanishes. The equation becomes

$$F_l = 20 \log_{10} \frac{j\omega L_m}{2Z_o} \qquad (7)$$

and can be applied to pair combinations with three and more separators.

Assuming for a moment that L_m does not change with frequency, the rms value of crosstalk loss at any frequency f_x can be computed from the rms value measured at one frequency f_o only, using the following expression derived from equation 7.

Table IX. Rms Far-End Crosstalk Coupling Loss Relative to Crosstalk Loss at 10 Kc Computed and Measured

Inductive Coupling, Pairs with Four Separators, in Db

Kc	Computed	Correction for L_m	Total Computed	Measured	Difference
10	0	0	0	0	0
50	−16.2	+0.8	−15.4	−15	0.4
100	−22.7	+1.5	−21.2	−21	0.2
200	−28.9	+1.9	−27.0	−27	0
400	−35.2	+2.1	−33.1	−33	0.1
800	−41.4	+2.1	−39.3	−39	0.3
1,000	−43.4	+2.1	−41.3	−41	0.2

Table X. Rms Far-End Crosstalk Relative to Crosstalk Loss at 50 Kc Computed and Measured for Pairs Coupled by Capacitance Unbalance and Mutual Inductance

Adjacent Pairs, in Db

Kc	Computed	Correction for L_n	Total Computed	Measured	Difference
50	0	0	0	0	0
100	− 6.3	+0.7	− 5.7	− 6	0.3
200	−12.6	+1.1	−11.5	−12	0.5
400	−18.9	+1.3	−17.6	−18	0.4
800	−25.1	+1.3	−23.8	−24	0.2
1,000	−27.1	+1.3	−25.8	−26	0.2

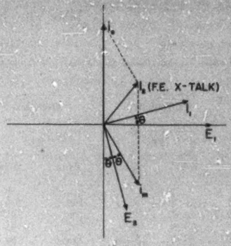


Fig. 18. Vector diagram of far-end crosstalk currents. The terminology used is the same as in Fig. 26. Voltage E_1 and current I_1 in the disturbing circuit induce electromagnetic and electrostatic components I_m and I_c of the crosstalk current I_f which flows toward the far end of the disturbed circuit

$$F_l \text{ at } f_o - F_l \text{ at } f_x = -20 \log_{10} \frac{f_x}{f_o}\frac{|Z_{fo}|}{|Z_{fx}|} \qquad (8)$$

where F_l at $f_o - F_l$ at f_x is the decrement in db of crosstalk coupling loss from the level F_l measured at the frequency f_o, Z_{fx} and Z_{fo} are characteristic impedances at frequencies f_x and f_o, where f_o and f_x, are in kc.

To the values so computed, a correction should be applied to take into account the change of L_m with frequency. As was shown by Hunter and Booth[8] mutual inductance, where affected by combined skin and proximity effects, is a complex quantity $L_m = L_{ma} + jL_{mb}$, which in cables of comparable size, decreases with frequency. L_{ma} drops at 300 kc to about 78% of the value measured at 10 kc and L_{mb}, insignificant at 10 kc, reaches its maximum of 13% of corresponding L_{ma} at 57 kc, decreasing at higher frequencies. Expressed in db, neglecting phase angle, this correction is shown in Table IX.

For example, equation 8 was used to compute the decrease of rms crosstalk coupling loss from rms value measured at 10 kc for pairs with four separators. The results are given in Table IX. It can be seen that after corrections for changes in L_m are introduced, the difference between measured and computed values is less than 0.5 db.

For pairs coupled by both capacitance and inductance this relation is more complex. At 1 kc the second term of equation 6 practically vanishes. As frequency increases, the resultant crosstalk currents will depend on the relative magnitudes of C_u and L_m and on the magnitude and angle θ of the characteristic impedance since the currents subtract vectorially as shown on Fig. 18.

To take into account all the factors, it

Table XI. Rms Near-End Crosstalk Coupling Loss Relative to Crosstalk Loss at 50 Kc Computed and Measured, in Db

Kc	Computed	Correction for L_n	Total Computed	Measured	Difference
Pairs Coupled by Mutual Inductance, Four Separators					
50	0	0	0	0	0
100	−5.7	+0.7	−5	−4	1.0
200	−11.0	+1.1	−9.9	−11.5	1.6
400	−16.0	+1.3	−14.7	−16.5	1.8
800	−20.7	+1.3	−19.4	−21.5	1.9
1,000	−22.1	+1.3	−20.8	−23.5	2.7
Pairs Coupled by Capacitance and Inductance, Adjacent Pairs					
50	0	0	0	0	
100	−5.7	+0.7	−5	−4.5	0.5
200	−11.0	+1.1	−9.9	−9.5	0.4
400	−16.0	+1.3	−14.7	−14.0	0.7
800	−20.7	+1.3	−19.4	−18.0	1.4
1,000	−22.1	+1.3	−20.8	−19.5	1.3

Table XII. Capacitance Unbalance in Unit 3, 3,000 Feet, 1 Kc

Type of Combination	Avg. μμf	Max. μμf	Rms μμf	Computed Rms Db Loss	Rms Near-End Crosstalk Loss Measured at 1 Kc, Db
Adjacent pairs	21.4	57.5	26.8	101.3	101.3
Alternate pairs	9.3	28.8	11.3	108.8	108.5
Two separators	4.4	22.4	6.8	113.2	112.9
Three separators	1.0	3.5	1.4	126.9	126.2
Four separators	0.1	1.0	0.3	140.6	137.4
Five separators	0.0	0.0	0.0		144.8
Six separators	0.0	0.0	0.0		133.7
Center to first layer	10.7	65.5	17.2	105.1	105.0
First to second layer	4.8	34.4	9.6	110.2	109.0
Alternate layers	5.3	25.8	7.6	112.3	110.7

is necessary to know the ratio of L_m to C_u, as will be shown in the section on capacitance unbalance. For an approximate computation, equation 8 can be used; it is advisable, however, to make the reference test of crosstalk losses at frequencies above 50 kc, at which Z_o and θ approach constant values as is indicated in Fig. 5 and inductive currents predominate. An example of the application of equation 8 to pairs having capacitive and inductive coupling is given in Table X. Measured values are taken from Fig. 15.

In the computation of changes with frequency of near-end rms crosstalk coupling loss, it should be taken into consideration that capacitive and inductive crosstalk currents tend to add instead of subtract and that they are attenuated on their way back to the near end.

It follows from equations 42 and 43 of Appendix V that near-end crosstalk coupling loss expressed in db is

$$N_l = 20 \log_{10} \left[\frac{j\omega C_u Z_o}{8} + \frac{j\omega L_m}{2Z_o} \right] \frac{1}{\sqrt{n}} \times \sqrt{\frac{1-e^{-4\alpha l}}{4\alpha\Delta}} \qquad (9)$$

where $\Delta n = l$, l is cable length in miles, and n is selected to be large enough so that phase shift and attenuation in the length Δ can be neglected ($\Delta = 20$–30 feet). N_l is near-end crosstalk coupling loss in length l in db, α is attenuation of line in nepers per mile, C_u and L_m are capacitance unbalance in farads and mutual inductance in henrys in the length l. Equation 9 can be transformed, assuming inductive couplings, in a similar way as equation 8 for computation of the decrease of coupling loss relative to the loss measured at a chosen frequency.

$$N_l \text{ at } f_o - N_l \text{ at } f_x = -20 \log_{10} \frac{f_x}{f_o} \frac{|Z_{fo}|}{|Z_{fx}|} \times \sqrt{\frac{\alpha_o}{\alpha_x}} \sqrt{\frac{1-e^{-4\alpha_o l}}{1-e^{-4\alpha_x l}}} \qquad (10)$$

where l is 0.57 miles and α_o and α_x are attenuations in nepers per mile at frequencies f_o and f_x. Other symbols are as given in equations 8 and 9.

Due to the waviness of near-end crosstalk loss curves, the prediction of changes with frequency of near-end crosstalk coupling loss is not as reliable as the prediction of far-end crosstalk. It will depend, first of all, on whether crosstalk has been measured at a frequency at which the curve of crosstalk is wavy, which can cause the reference point, that is, the measured value (see Fig. 16) to shift considerably from its true position. Most favorable conditions will be established if the frequency at which crosstalk is measured is chosen low enough so that the cable is not longer than 1/4 wavelength (to avoid waviness of curves) and high enough to make inductive currents predominate; 50–70 kc seem to be appropriate for this test.

Using equation 10, changes of rms coupling losses were computed from the measured at 50 kc and compared with changes of crosstalk as given in Fig. 16, with the waviness of the curves smoothed out. The data are given in Table XI.

It can be seen that the deviation of the computed values from the measured is greater for near-end than for far-end crosstalk.

It appears from the quoted examples that crosstalk coupling loss can be predicted for the whole carrier frequency range from a test performed at only one frequency, with good accuracy at far-end and with reasonable approximation at near end.

CAPACITANCE UNBALANCE AS AN INDEX OF CROSSTALK

It is common to measure capacitance unbalance in order to determine the value of crosstalk within the voice-frequency range. It will be useful to analyze whether this test can be regarded as a reliable index of inductance unbalance and, by it, an index of crosstalk at carrier frequency as well.

At voice frequency (1 kc), the rms capacitive crosstalk currents between adjacent pairs are ten times greater than the rms inductive currents. Since these currents add or subtract vectorially at an angle of about 86 degrees, the resultant vector differs from capacitive current by a fraction of 1% only. The inductive component of crosstalk currents can be, therefore, neglected and it is enough to perform a simple test of capacitance unbalance to determine crosstalk characteristic of a cable at voice frequency.

The values of capacitance unbalance measured between pairs of unit 3 are given in Table XII. The crosstalk coupling loss at 1 kc was computed from rms values of capacitance unbalance employing equation 6 and neglecting the term for mutual inductance. For comparison, values of near-end crosstalk measured at 1 kc are given in the same table.

It can be seen that for the combinations of pairs having capacitive couplings, the rms values of crosstalk are practically the same, whether measured as capacitance unbalance or as crosstalk coupling loss. For pairs with many separators, the relation does not hold since inductive coupling alone is causing crosstalk, the absolute value of which is, however, negligible in the order of a fraction of one crosstalk unit (over 120 db).

Table XIII. Equivalent Mutual Inductance, Factor $m = \frac{L_m}{C_u} \frac{\mu\mu h}{\mu\mu f}$

	Pairs Combination		Two Separators
	Adjacent	Alternate	
Unit 1	6.1	5.7	9.4
Unit 2	4.7	6.1	9.1
Unit 3	6.0	6.9	7.5
Unit 4	6.5	6.1	7.4
Grand Rms, four units	5.9	6.3	8.0
Two cables spliced, unit 3	5.4	6.8	7.7

Note: Both L_m and m decrease with frequency by factors 0.90, 0.83, 0.82, and 0.78 at 50, 100, 150, and from 300 kc up, respectively.

As frequency is increased, the inductive components of crosstalk increase rapidly. At 10 kc (for adjacent pairs) the magnitudes of capacitive and inductive currents become equal and at 1,000 kc inductive currents are roughly twice as high as capacitive. For nonadjacent pairs, the ratio increases in favor of inductive currents. Direct tests of crosstalk or of mutual inductance are relatively complicated, and a method was sought to correlate capacitance unbalance with mutual inductance so that by testing one quantity, sufficient information can be derived to estimate another.

It was found that such correlation does exist in cables in which conductors are insulated with accurately applied polyethylene. In rms values, there is about 6 mμh (millimicrohenry) of mutual inductance per 1 μμf (micromicrofarad) of capacitance unbalance between adjacent pairs. This can be understood considering the source of unbalances. If pairs were not twisted then a very strong inductive and capacitive coupling would exist between them due to their proximity in the cable. By transposition (pair twisting) those strong couplings are changed into a multitude of small couplings of opposite signs, which mutually almost cancel, and the minute residues left are called capacitance unbalance and mutual inductance. These residues amount to a fraction of 1% of the original inductive and capacitive couplings. If twisting of pairs is inaccurate, both unbalances will be affected.

Based on this premise, an rms "equivalent" mutual inductance has been computed from values of rms crosstalk loss measured at 1 kc and 1,000 kc using equation 11, assuming that L_m is zero at 1 kc and $\theta = 0$ at 1,000 kc. Since mutual inductance L_m changes with frequency, as previously stated, the values of L_m computed from crosstalk at 1,000 kc have been corrected so that the values of factor m given in Table XIII refer to an equivalent L_m at 1 kc.

A correlation can be seen between rms values of capacitance unbalance and mutual inductance. For a particular pair combination the ratio m, that is, the number of units of mutual inductance per one unit of capacitance unbalance, varies not more than 20% or, in crosstalk inductive currents ±1.9 db from the grand rms of the particular type of pair combinations. As the separation between pairs increases the ratio of L_m/C_u increases also, showing the shielding effect of intervening pairs by relative reduction of capacitance unbalance C_u.

Once the ratio between the equivalent L_m and C_u is established, the crosstalk coupling loss can be computed from equation 11 substituting mC_u for L_m. This is applicable for pairs with measurable capacitance unbalance (adjacent, alternate, and pairs with two separators). For other pair combinations, crosstalk loss has to be estimated from the cable design and quality of pair twisting, indicated by the capacitance unbalance test. The applicable equations are given in the following. The rms far-end crosstalk current ratio expressed in crosstalk units (CU) is

$$F_{cu}' = \frac{j\omega C_u Z_o}{8 \times 10^6} - \frac{j\omega L_m}{2Z_o \times 10^3}$$

$$= \frac{j\omega C_u}{8 \times 10^6}\left[Z_o - \frac{4m10^3}{Z_o}\right] = C_u \times P_m \quad (11)$$

where

F_{cu}' = rms crosstalk current ratio, C_u
L_m = rms mutual inductance, mμh, at frequency of test
C_u = rms capacitance unbalance, μμf
Z_o = characteristic impedance, ohms
m and corrections for frequency are given in Table XIII

The value of P_m has been computed for various values of m for 19 Awg 0.083 cables with the two terms in parenthesis of equation 11 added vectorially and corrections introduced for changes of L_m with frequency. Values of P_m are given in Fig. 19.

Equation 11 can be used in conjunction with Table XIII and Fig. 19 to compute rms far-end crosstalk coupling loss at any frequency when the value of m is known and capacitance unbalance C_u is measured. Expressed in db, far-end crosstalk coupling loss is equal to $20 \log_{10} F_{cu}' - 120$.

Near-end crosstalk current ratio in CU can be computed from equation 11 by changing the minus sign to plus and allowing for attenuation at the near end; however, due to the nature of near-end crosstalk, the estimate will be less accurate than the estimate of far-end crosstalk.

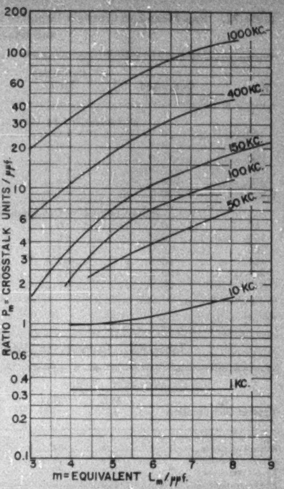


Fig. 19. Crosstalk units per 1 μμf of capacitance unbalance for various values of m (equation 11)

The method discussed for approximating crosstalk from the measured capacitance unbalance is not suggested as a substitute for direct measurements of crosstalk. It is intended only to draw attention to the fact that capacitance unbalance in a general way is an index of symmetry and crosstalk characteristics of cable at carrier frequencies and to show that rms crosstalk coupling loss in a range of 1–1,000 kc can be estimated from capacitance unbalance with an accuracy of a few db, after the values of m are established for a particular cable design.

The ratio $L_m/C_u = m$, given in Table XIII, has been established for the design and manufacturing methods employed to produce the polyethylene-insulated cable tested. If the design is changed, for instance, regularity of extrusion of conductors, type of twisting, number of pair twists, and so on, then the ratio of equivalent $L_m/C_u = m$ will change also.

The rms value of m for any design of 19 Awg 0.083 cable can be established using expression 11 and Fig. 19 from a test of capacitance unbalance at 1 kc and of cross-talk coupling loss at any convenient carrier frequency.

It is expected that for 22 Awg 0.083 cables the same relation will be valid since mutual capacitance and inductance of circuits are practically the same as for 19 Awg 0.083 type of cable. For 19 Awg 0.066 cables, the ratio L_m/C_u should increase due to the reduction in mutual capacitance and increase in inductance of circuits. The absolute value of equivalent mutual inductance should not increase, however, by more than 1.5 db since the absolute value of inductance of circuits is increased by 20% but the separation between pairs is also increased. Inductive crosstalk currents will not increase, however, due to compensating increase in Z_0.

SUMMATION OF CROSSTALK WITH LENGTH

To check whether the rule of random summation of currents with increased length of cable will hold when cables are straight-spliced, two reel lengths were spliced and tested. On purpose, most unfavorable conditions were arranged. Spliced units were originally paired, stranded, and cabled in one continuous run. If, in the process of manufacture, some of the elements (diameter of conductors, uneven twisting) were systematic then this should show up in the crosstalk tests on pairs spliced color to color. The results of the test are given in Table XIV.

It can be seen that the levels of crosstalk loss were different in the two reels. When rms values of crosstalk currents were added $(F_{12}'=\sqrt{F_1'+F_2'})$, it was found that for the adjacent pairs, that is, for the controlling pair combinations, the measured value was somewhat better than would have been expected from the random summation of currents. A deviation of the same magnitude but of an opposite sign can be seen for far separated pairs between which the crosstalk loss is roughly 20 db higher than between adjacent pairs. The deviations were small, within possible test error, particularly at 1,000 kc. No indications were found in this limited test that the rule of random summation of currents may not apply for straight-spliced cables.

Applying the principle of random summation, the rms values of crosstalk coupling loss given in the tables and graphs for a 3,000-foot length of cable can be converted to other lengths using equations 41 and 43 of Appendix V.

FAR-END CROSSTALK

$$F_l=(F_o-2.4)-10\log_{10} n \qquad (12)$$

where

F_l = crosstalk in db for chosen characteristic (pairs combinations, unit or cable) in n miles of cable
F_o = crosstalk in db read on Tables II–VI or Figs. 15(A) and 15(B) for 3,000-foot length of cable (−2.4 db is added to convert crosstalk F_o into crosstalk in a mile length)
n = miles

NEAR-END CROSSTALK

$$N_l=N_o-20\log_{10}\sqrt{\frac{1-e^{-4\alpha l}}{1-e^{-4\alpha l_0}}}=N_o-D_o \qquad (13)$$

where

N_o = value of crosstalk in a 3,000-foot length of cable given in Figs. 16(A) and 16(B) and Tables VII and VIII
α = attenuation, nepers per mile
l = length, miles
l_0 = 0.57 mile (3,000 feet)
N_l = near-end crosstalk in length l, db

The second term of equation 13, D_o, is given in Fig. 20.

Shield Parameters

The Alpeth sheath of the plastic telephone cable, in addition to providing a protective outer covering for mechanical and handling purposes, also serves to reduce interference from power lines and interference from atmospheric discharges at carrier frequencies. The sheath also provides a low-resistance path so that lightning currents in the vicinity do not build up excessive voltages along the shield and cause dielectric breakdown of the underlying insulation. It will be indicated later that it is desirable to have a low shield resistance (or impedance).

One type of voltage, induced in the cable at voice frequencies from external sources, is due to magnetic induction and is proportional to the magnitude of the harmonics of the power currents in the disturbing circuit. This voltage is induced longitudinally along the cable and is reduced to zero if the shield resistance is zero (see Appendix VI) and the shield is grounded at both ends through ground connections having zero resistance. This condition, of course, is never achieved in practice because the shield and the ground connections both have resistance. Because the two wires forming a telephone pair are twisted and are close together, practically no differential voltage is induced directly between the wires themselves to produce noise currents. However, there is an induced voltage between the two wires of a pair and ground, or along the two wires in parallel, which causes noise current in the pairs due to cable pair unbalances.

The second type of induced voltage is proportional to the voltages of the power circuit and is called electric induction. This type of induction appears between the core and ground if the shield is floating and may be easily eliminated by grounding the shield at sufficiently short intervals. The discussion of shield factors of the cable assumes that this type of induction has been eliminated by proper grounding.

SHIELD FACTOR FOR MAGNETIC INDUCTION

The shield factor may be defined as the ratio of the induced voltage along the cable core when the shield is present (assuming an electrically short length with perfect grounds at both ends) to the induced voltage when the shield is absent. Good shielding is obtained when the shield factor is low. The shield factor as explained in Appendix VI may be expressed as follows:

$$S_f=\frac{Z_{11}-Z_{12}}{Z_{11}} \qquad (14)$$

where

Table XIV. Change of Far-End Rms Crosstalk Loss With Increased Length, Unit 3, Db

Kc	Reel 1	Reel 2	Measured, Reels 1 and 2	Computed, Reels 1 and 2	Difference
			Adjacent Pairs		
1	100.8	97.0	95.5	95.5	0
10	90.9	87.7	86.9	86.0	+0.9
150	71.1	71.5	69.5	68.3	+0.8
400	61.0	63.4	60.3	59.1	+1.2
1,000	54.2	55.1	50.5	51.6	−1.1
			Four Separators		
1	136.8	142.9	135.6	135.9	−0.3
10	113.2	114.1	110.5	110.6	−0.1
150	88.4	93.2	86.3	87.1	−0.8
400	80.5	84.0	77.0	78.9	−1.9
1,000	72.3	74.2	66.6	68.8	−2.2

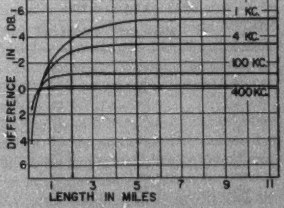


Fig. 20. Near-end crosstalk coupling loss versus length of cable (equation 13)

Table XV. Shield Factor*

Frequency, Kc	ρ_g† = 10 Measured	ρ_g† = 10 Computed	ρ_g = 100 Measured	ρ_g = 100 Computed	ρ_g = 1,000 Measured	ρ_g = 1,000 Computed
Folded Shield						
1	0.148	0.140	0.129	0.123	0.115	0.109
2	0.0773	0.0730	0.0683	0.0646	0.0607	0.0574
4	0.0411	0.0382	0.0366	0.0340	0.0321	0.0300
8	0.0219	0.0199	0.0195	0.0177	0.0171	0.0155
10	0.0177	0.0161	0.0158	0.0144	0.0138	0.0125
20	0.0092	0.0083	0.0081	0.0074	0.0072	0.0065
40	0.0047	0.0043	0.0042	0.0038	0.0037	0.0034
80	0.0022	0.0022	0.0020	0.0020	0.0019	0.0018
100	0.0018	0.0017	0.0016	0.0016	0.0015	0.0014
200	0.0009	0.0009	0.0008	0.0008	0.0007	0.0007
400	0.0004	0.0004	0.0004	0.0004	0.0003	0.0003
800	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001
1,000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
Helical Shield‡						
1	0.488	0.470	0.442	0.424	0.405	0.386
2	0.284	0.269	0.256	0.241	0.231	0.215
4	0.172	0.146	0.155	0.130	0.138	0.114
8	0.119	0.0768	0.107	0.0683	0.0956	0.0600
10	0.110	0.0621	0.0997	0.0555	0.0882	0.0485
20	0.0968	0.0322	0.0867	0.0286	0.0774	0.0253
40	0.0918	0.0166	0.0830	0.0148	0.0742	0.0132
80	0.0912	0.0084	0.0833	0.0077	0.0765	0.0070
100	0.0905	0.0068	0.0828	0.0061	0.0743	0.0055
200	0.0827	0.0034	0.0769	0.0032	0.0684	0.0028
400	0.0723	0.0016	0.0679	0.0015	0.0611	0.0013
800	0.0596	0.0006	0.0569	0.0006	0.0521	0.0005
1,000	0.0563	0.0004	0.0542	0.0004	0.0496	0.0004

* For cable 20 feet above ground and computed values are based on equations 14 to 17 inclusive.

† ρ_g is the earth resistivity in meter-ohms.

‡ Computed values for the helical-shield are made on the basis that the shield is applied as a fold having the same d-c resistance as the helical tape shield for which measured values are given.

S_f = shield factor
Z_{11} = ground return self-impedance of the shield, ohms per mile
Z_{12} = ground return mutual impedance between the shield and the core of the cable, in ohms per mile

The term $(Z_{11}-Z_{12})$ is the leakage impedance between the shield and the shielded conductors. It also is called the transfer imedance or surface transfer impedance in ohms per mile and is designated as Z_{ab}. Therefore, equation 14 may be rewritten as:

$$S_f=\frac{Z_{ab}}{Z_{11}} \tag{14A}$$

The calculation of shield factor is not practical at higher frequencies since it assumes grounds at intervals which are short as compared with the wavelength. For example, at 300 kc, grounds 100 feet apart would have an electrical length of about 1/20 wavelength (18 degrees) which is about the maximum length for an electrically short line. Computed and measured shield factors are tabulated in Table XV for a cable having a folded aluminum shield with a measured and calculated d-c resistance of 2.25 and 2.26 ohms per mile, respectively, and a cable having the same core but having two helically applied aluminum tapes 1.5 inches by 0.005 inch with a 2.6-inch pitch and a measured and calculated d-c resistance of 8.79 and 8.74 ohms per mile, respectively. The d-c resistance of the folded shield was calculated as if it were a tube and the helical tape was calculated assuming no contact between turns. The diameter under the shield for each cable was 1.30 inches. In order to determine the shield factor it is necessary to calculate first the surface transfer impedance and next the ground return self-impedance of the shield.

SURFACE TRANSFER IMPEDANCE OF THE SHIELD

The surface transfer impedance (Z_{ab}) which appears in equation 14(A) is the ratio of the voltage induced along the inside of the shield to the current flowing in the shield. The values are usually converted to volts per mile for one ampere of disturbing current flowing in the shield. It is helpful to present the data in decibels above or below one ohm per mile (i.e. 20 $\log_{10}|Z_{ab}|$ where $|Z_{ab}|$ is ohms per mile). The magnitude of the surface transfer impedance for a shield consisting of a tube may be calculated from the following formula developed by Schelkunoff:[9]

$$|Z_{ab}|=R_{dc}'\left[\frac{u}{\sqrt{\cosh u-\cos u}}\right] \tag{15}$$

where

R_{dc}' = shield d-c resistance, ohms per mile
$u=18.97t_s\sqrt{f}$
t_s = thickness of the shield, inches
f = frequency, in kc, for aluminum tubes

If the shield consists of helically applied tapes, this formula does not hold because the currents spiral around the core. This is particularly true if the shielding tape is made of aluminum because of the high contact resistance between the surfaces of the aluminum tape at their overlaps.

The surface transfer impedance was measured using a method similar to that described by Gooding and Slade.[10] The conductors were all connected and served

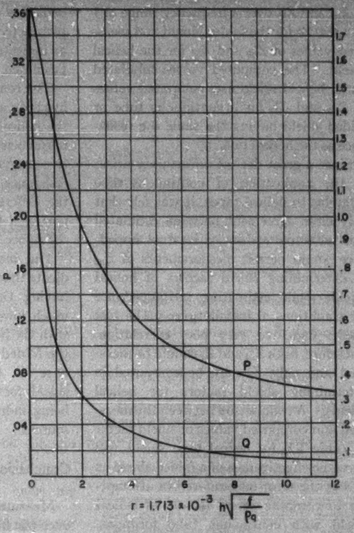
Table XVI. Surface Transfer Impedance (Z_{ab}), Internal Impedance (Z_i), and Shield to Core Impedance (Z_o')

Frequency, Kc	Folded Shield: Z_{ab}, Db from 1 Ohm per Mile, Measured	Folded Shield: Z_{ab}, Db from 1 Ohm per Mile, Computed	Folded Shield: Z_i, Ohm per Mile, Measured*	Folded Shield: Z_i, Ohm per Mile, Computed	Folded Shield: Z_o', Ohm	Helical Shield: Z_{ab}, Db from 1 Ohm per Mile	Helical Shield: Z_i, Ohm per Mile	Helical Shield: Z_o', Ohm
1	7.46	7.05	2.14+j0	2.25+j0.009	15 +j11.5	19.9	8.76+j 2.0	23 +j17.2
2	7.54	7.05	2.18+j0	2.25+j0.017	14.4+j 7.6	20.2	8.78+j 4.0	18.6+j 4.0
4	7.68	7.05	2.23+j0	2.25+j0.034	10.6+j 4.7	21.2	8.78+j 7.2	17.7+j 6.9
8	7.89	7.05	2.32+j0	2.25+j0.069	9.5+j 2.9	23.5	8.1 +j 14	15.5+j 4.3
10	7.89	7.05	2.37+j0	2.25+j0.086	9.2+j 2.4	24.7	8.0 +j 16	15.2+j 3.9
20	7.89	7.05	2.37+j0.20	2.25+j0.172	8.9+j 1.4	29.3	8.8 +j 29	14.1+j 3.4
40	7.89	7.02	2.37+j0.35	2.25+j0.345	8.8+j 0.98	34.5	24 +j 51	13.1+j 3.4
80	7.24	6.98	2.44+j0.70	2.25+j0.69	8.7+j 0.72	40.0	91 +j 60	11.4+j 3.4
100	7.24	6.86	2.45+j0.89	2.32+j0.86	8.6+j 0.63	41.7	115 +j 62	10.8+j 3.3
200	7.08	6.82	2.56+j1.36	2.59+j1.64	8.3+j 0.45	46.8	206 +j137	9.9+j 2.7
400	6.02	5.94	3.79+j1.95	3.15+j2.80	8.2+j 0.34	51.6	409 +j222	8.8+j 2.2
800	4.08	3.92	4.47+j4.12	4.40+j4.00	7.8+j 0.30	55.8	588 +j548	8.6+j 2.4
1,000	2.92	2.80	4.69+j4.44	5.28+j5.27	7.5+j 0.28	57.3	709 +j616	8.1+j 2.2

The reactive component was too small to measure on available equipment at frequencies below 20 kc for the folded shield; consequently, this component is tabulated as zero.

as the return path. (A diagram is shown in Fig. 23(A) which gives the measuring circuit employed.) A coaxial cable having a solid aluminum sheath was used to establish the accuracy of the circuit because such a configuration permits Z_{ab} to be accurately calculated using equation 15. The results obtained for folded and helical shields are shown in Table XVI, columns 2 and 7, for measurements over a frequency range of 1 kc to 1,000 kc. Computed values of Z_{ab} employing equation 15 with computed d-c resistance for the folded shield cable are also tabulated in Table XVI, column 3. The tabulated results clearly show that the computed and measured values of surfaces transfer impedance for a folded shield are in close agreement. This is expected since a folded shield behaves as a continuous tube for normal transmission. However, the helical shield deviated markedly from the computed values. These computed values are not tabulated.

Fig. 21. Functions P and Q for use in equation 16



GROUND RETURN SELF-IMPEDANCE OF THE SHIELD

The ground return self-impedance of the shield as shown in equation 14(A) is equal to the sum of an internal impedance (Z_i) and an external impedance (Z_e). The external impedance (Z_e) includes the effect of the magnetic flux in the earth and between the surface of the shield (if aerial) and the surface of the earth and is, therefore, the external series self-impedance of a tube or round conductor above earth with perfect ground connections. Z_e is quite difficult to measure and is therefore estimated. Z_i is the internal or intrinsic impedance of a tube with an external return path and is sometimes referred to as Z_{bb} and is nearly equal to Z_{aa} for large diameters and thin shields. The latter impedance (Z_{aa}) is the internal impedance of a tube with an internal return path. Since it is simpler to measure, it was measured instead of Z_{bb}.

The internal impedance (Z_{aa}) was measured employing a method similar to that used for measuring Z_{ab} except that both the phase and magnitude of the current and voltage along the outside of the shield were measured instead of magnitudes only of the current on the outside of the shield and voltage on the inside of the shield. It is necessary to know the phase angle of the internal impedance in order to compute the denominator of the shield factor. (A diagram of the test circuit is shown in Fig. 23(B).) Since there are innumerable conditions in the field (i.e., ground resistivity, cable height, etc.), it was felt that the computed external impedance for a fixed cable height of 20 feet and several ground resistivities would approximate field conditions. This may be derived from the following equations due to Carson[11] for Z_{11} and to Schelkunoff[9] for Z_i:

$$Z_{11}=Z_i+j7.41\omega\times10^{-4}\log_{10}24h/b+6.44\omega(P+jQ)\times10^{-4}\text{ ohms per mile}\qquad(16)$$

where

$$Z_e=j7.41\omega\times10^{-4}\log_{10}24\,h/b+6.44\omega(P+jQ)\times10^{-4}\qquad(16\text{A})$$

where Z_i (or Z_{bb}) is given by:

$$Z_{bb}=Z_i=R_{dc}'\left[\frac{u}{2}\frac{(\sinh u+\sin u)+j(\sinh u-\sin u)}{\cosh u-\cos u}+\frac{t_s}{2b}\right]\text{ ohms per mile}\qquad(17)$$

where

h = distance of shield above ground, feet
b = outside radius of shield, inches
ρ = earth resistivity, meter-ohms
t_s, u, f = same as given in equation 15
$\omega=2\pi f$
P, Q = special functions given in Fig. 21 for argument $r=1.713\times10^{-3}h\sqrt{f/\rho_g}$

The first two terms on the right-hand side of equation 16 represent the series impedance of a shield with ground return if the ground is a perfect conductor. The last term represents the effect of finite conductivity of the ground. The last two terms are the external impedance (Z_e) which is due to fluxes caused by current in the shield and return current in the earth. This flux exists in the earth and between the surface of the earth and surface of the shield. The magnetic flux pattern is transverse to the axis in cylinders and helical in helices. The external transverse component of the helical flux is much more important than the external axial component and, therefore, the external field (per ampere) is much the same for cylindrical shields and for practical forms of helical shields. The external part (Z_e) of the self-impedance (Z_{11}) may be calculated from equation 16(A) for a cylindrical shield and estimated for a helical shield. The height of 20 feet mentioned is not critical since, in the practical case, the image conductor for computing Z_e is quite deep and Z_e changes slowly with the cable height. The calculated values of Z_e for ρ_g of 10 are 1.30+j15.6 at 1 kc and 284+j12,250 at 1,000 kc; these compare to values of internal impedance (Z_i) of 8.76+j2.00 and 709+j616 for the helical tape shield and internal impedance of 2.14+j0 and 4.69+j4.44 for the folded tape shield, respectively. All values are given in ohms per mile.

Once the self-impedance (Z_{11}) is determined for a given ground resistivity, the shield factor may then be determined for this value using equation 14. Table XV gives shield factors determined by computing the external impedance for ρ_g equal to 10, 100, and 1,000 meter-ohms, and both measured and computed values of Z_{ab} and Z_{bb} over a frequency range of 1 to 1,000 kc for the folded shield and measured values of Z_{ab} and Z_{bb} for the helical shield. The computed values tabulated under helical shield uses computed values of Z_{ab} and Z_{bb} for a continuous tube or folded shield having the same d-c resistance as the helical tape.

Since a great deal of uncertainty arises in the application of continuous tube formulas to helical tapes, it was felt that only close agreement between estimated and measured values of internal impedance would permit measurements of Z_{ab} in determining shield factors of helical shields while estimating self-impedance for conditions of installations. The test results indicated very poor correlation. Therefore, both Z_{ab} and Z_i should be measured and only Z_e may be computed in determining shield factors for helical shields. A comparison between the measured and computed values of shield factor in Table XV for helical tape shield will show the effectiveness of a folded shield as well as the error encountered in attempting to compute shield factors of helical shield with continuous tube formulas. In using the data presented in this table the relative shield factors are applicable only to electrically short lines. As previously indicated, the comparisons shown are not valid unless this is the case and may be somewhat different in actual cases where grounding of the shield is not frequent enough to ground at electrically short distances.

Shield-Core Impedance (Surge Impedance)

The shield to core impedance of the cable (between the shield and all the pairs) was measured in order to determine the impedance offered to lightning surges. Atmospheric discharges, i.e., lightning currents, in the vicinity of the cable cause currents to flow along the shield and if the resistance of the shield is low the voltage drop along the shield will be low. The longitudinal current along the shield produces a voltage inside the shield which in extreme cases, may cause breakdown to core and in other cases cause noise currents to flow along the cable pairs if there is an unbalance to ground which is normally the case. The amount of current that will flow will be determined, therefore, by this shield to core impedance of the cable and the unbalance.

In order to measure this impedance, the cable was regarded as a coaxial cable having all pairs connected as an inner conductor and the shield as outer conductor. The impedance was determined by measuring the short- and open-circuit impedances on the same bridge as used for parameter measurements and calculated as indicated previously. The results are tabulated in Table XVI.

The results show that the folded shield has only a slightly lower impedance than the helically applied shield at the higher frequencies. This is because the inductive reactance is much lower than the a-c resistance. However, at low frequencies, the difference in the impedance is proportional to the square root of the shield resistance.

The measurements on the shield indicate that the cable with the folded shield having the lower d-c resistance has a much lower shield factor than the cable with the helical shield; this is true even if the folded shield has the same d-c resistance as the helical shield. This low shield factor will result in less voltage being induced in the core from external sources.

Conclusions

Measurements have been conducted over the frequency range of 1 kc to 1,000 kc and a temperature range of 0 F to 140 F to establish the primary and secondary parameters of a typical multipair polyethylene-insulated Alpeth-sheathed telephone cable. Crosstalk properties over the same frequency range have been measured. Formulas have been suggested which permit calculation of the parameters from cable dimensions with a reasonable degree of accuracy. The extent of the effects of immersion of the cable core in water or of moisture contamination on the parameters are presented.

It is indicated that from the measurements of crosstalk at one frequency the crosstalk at other carrier frequencies can be estimated with good accuracy for far-end and with fair accuracy for near-end crosstalk coupling loss. A correlation between capacitance unbalance and crosstalk at higher frequencies for pairs having capacitive coupling is indicated and the ratio between capacitance unbalance and equivalent mutual inductance given. The change of crosstalk coupling loss as a function of cable length has been explored and a method of computation indicated.

The shield parameters have been measured at room temperature over the frequency range of 1 kc to 1,000 kc and these indicate the advantages of using a cable with a folded or tubular low-resistance metallic shield.

Appendix I. Test Equipment

The primary parameters were measured using a Siemens and Halske impedance-measuring bridge, type *Rel 3 R 217 P(n)*. The open-circuit admittance and the short-circuit impedance of the cable pair under test were measured at the frequency and temperature applicable. Such measurements were made at frequencies from 1 to 1,000 kc at temperatures from 0 F to 140 F in steps of 20 degrees. For measuring the open-circuit admittance the bridge is connected as a Wheatstone bridge modified for a-c applications as devised by Wien (specimen and standard in adjacent bridge arms); for measuring the short-circuit impedance the bridge is connected as a Maxwell bridge (specimen and standard in opposite arms). Various checks against laboratory standards were made with the bridge and it was found to be accurate within the following manufacturers' guaranteed values: for open-circuit admittance the accuracy is ±1% for capacitance and ±1% for conductance for frequencies below 700 kc and ±2% for conductance for frequencies above 700 kc; for short-circuit impedance the accuracy is ±1% for both resistance and inductance.

The oscillator used in conjunction with the bridge was a General Radio Standard Signal Generator type *1001-A* in the frequency range 400 kc to 1,000 kc and a Hewlett-Packard Audio Frequency Oscillator

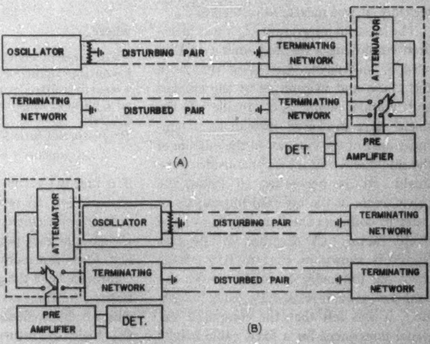


Fig. 22. Schematic diagram of circuits used in measuring crosstalk coupling loss

A—At far end
B—At near end

Vol. 7
Pt. 1
1959

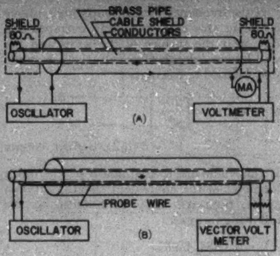


Fig. 23. Schematic diagram for circuit used in measuring the surface transfer impedance (Z_{ab}) of the shield (A) and the internal impedance (Z_i) of the shield (B)

model *233A* in the frequency range 1 kc to 400 kc. Two detectors were employed: the Rhode and Schwarz Tunable indicating amplifier type *UBM* and the Hewlett-Packard Vacuum Tube Voltmeter model *400 H*.

Crosstalk was measured by a comparison method. The voltages in the disturbed and disturbing circuits were compared using a standard attenuator connected to the disturbing line. In measuring, the attenuator was adjusted until the same voltages were read by the detector when switched from the disturbed circuit to the attenuator. This eliminated the need for calibration or for corrections of readings of the disturbing and disturbed voltage levels. To provide terminations for phantom and other tertiary circuits, the center taps of terminations on all pairs in the cable were grounded as well as the center tap of the balanced attenuator. The oscillator, preamplifier, and detector were of the balanced type, not grounded at the center tap. The schematic diagram of the connections for measurement of far-end output to output crosstalk is given in Fig. 22(A) and for measurement of crosstalk at the near end is given in Fig. 22(B).

As shown in the circuit diagrams (Fig. 22), one end of the disturbing pair and both ends of the disturbed pair were terminated into precision networks. The maximum deviation of impedances of these networks (real and imaginary) from corresponding characteristic impedance of a typical pair at frequencies between 1 kc and 1,000 kc was less than ±2%. All other pairs were terminated in pure resistances.

The attenuator used was type *Rel 3D 24*, Siemens and Halske, with a range of 160 db at frequencies 1–20 kc and 130 db at 20–1,000 kc. Checked against laboratory standards, the accuracy of the attenuator was established to be within 0.2 db. The oscillator, Siemens and Halske *Rel W 317a*, has a maximum frequency error of ±0.2% up to 100 kc and 0.5% at frequencies above 100 kc. Preamplifier and superheterodyne receiver, Siemens and Halske *Rel 3U 42* and *Rel 3U 412a*, were sensitive to 0.3-microvolt signals and the sensitivity was used when extraneous noise voltages were filtered by using 10-cps (cycles per second) or 100-cps bandwidths when measuring high attenuations (above 120 db). At frequencies above 300 kc and losses below 75 db, for convenience a vacuum-tube voltmeter, Hewlett and Packard type *400H*, was used as a detector.

The measuring circuits employed for shielding studies are shown in Fig. 23.

Appendix. II Cable Parameter Relations

The electrical characteristics of a balanced cable pair are given by the following well-known[13] relations:

$$\gamma=\alpha+j\beta=\sqrt{ZY} \text{ propagation coefficient} \quad (18)$$

$$Z_0=\sqrt{Z/Y}=\sqrt{\frac{R+j\omega L}{G+j\omega C}}$$

$$= \text{characteristic impedance} \quad (19)$$

$$Z_0\gamma=Z=R+j\omega L \quad (20)$$

$$\gamma/Z_0=Y=G+j\omega C \quad (21)$$

where

α=attenuation, nepers per mile (db per mile=8.686 X nepers per mile)
β=phase shift, radians per mile
$Z=R+j\omega L$=series impedance, ohms per mile
$Y=G+j\omega C$=parallel admittance, mhos per mile
R=series resistance, ohms per loop mile
L=series inductance, henrys per loop mile
G=parallel conductance, mhos per mile
C=mutual capacitance, farads per mile
$\omega=2\pi f'$ where f' is the frequency in cps

In equation 18, $\gamma=\alpha+j\beta$, it may be shown[13] that in terms of the distributed parameters R, L, G, and C:

$$\alpha=\sqrt[4]{(RG-\omega^2LC)^2+(G\omega L+R\omega C)^2}\times \cos\frac{1}{2}\left[\tan^{-1}\frac{G\omega L+R\omega C}{RG-\omega^2LC}\right] \text{ nepers per mile} \quad (22)$$

$$\beta=\sqrt[4]{(RG-\omega^2LC)^2+(G\omega L+R\omega C)^2}\times \sin\frac{1}{2}\left[\tan^{-1}\frac{G\omega L+R\omega C}{RG-\omega^2LC}\right] \text{ radians per mile} \quad (23)$$

At high frequencies equations 22 and 23 may be expressed as:

$$\alpha\cong\frac{1}{2}G\sqrt{L/C}+\frac{1}{2}R\sqrt{C/L} \quad (22A)$$

$$\beta\cong\omega\sqrt{LC} \quad (23A)$$

At low frequencies equations 22 and 23 may be expressed as:

$$\alpha\cong\sqrt{\omega RC/2} \quad (22B)$$

$$\beta\cong\sqrt{\omega RC/2} \quad (23B)$$

In addition to the foregoing, the following well-known relations[13] are useful; these are given in terms of the series and parallel measured parameters for length, l, of the line:

$$Z_{sc}'=R'+j\omega L'=Z_0\tanh\gamma l$$

$$Y_{oc}'=1/Z_{oc}'=G'+j\omega C' =\frac{1}{Z_0\coth\gamma l}=\frac{\tanh\gamma l}{Z_0}$$

Therefore:

$$Z_0=\sqrt{Z_{oc}'Z_{sc}'}=\sqrt{Z_{sc}'/Y_{oc}'} \quad (24)$$

and

$$\tanh\gamma l=\sqrt{Z_{sc}'/Z_{oc}'}=\sqrt{Z_{sc}'Y_{oc}'} \quad (25)$$

where

Z_{sc}'=short-circuit impedance of the pair in ohms for length l
R'=series loop resistance in ohms for length l
L'=series loop inductance in henrys for length l
Y_{oc}'=open-circuit admittance of the pair in mhos for length l
G'=mutual conductance in mhos for length l
C'=mutual capacitance in farads for length l
l=length of line in miles under test

If the short-circuit impedance and open-circuit admittance are measured then the characteristic impedance Z_0 may be calculated from equation 24 and $\tanh \gamma l$ may be calculated from equation 25. Further, if then Z_0 and $\tanh \gamma l$ are known it is merely necessary to determine γ from $\tanh \gamma l$ and then, using equations 20 and 21, determine the primary parameters, R, L, C, and G.

In order to obtain γ, α and β are calculated from the following relationships,[13] letting $\tanh\gamma l=A+jB=D\underline{/\phi}$

$$\tanh 2\alpha l=\frac{2A}{1+A^2+B^2} \quad (26)$$

$$\tan 2\beta l=\frac{2B}{1-(A^2+B^2)} \quad (27)$$

where

$$\tanh\gamma l=D\underline{/\phi} =\sqrt{(R'+j\omega L')(G'+j\omega C')} \quad (28)$$

$$A=D\cos\phi$$

$$B=D\sin\phi$$

From equations 26 and 27 the real functions of the components of γ can be determined and α and β can then be determined from the tables of these real functions and then adjusted to a per-mile basis. This gives γ on a per-mile basis which then is substituted in equations 20 and 21 yielding R, L, C, and G on a per-mile basis.

Appendix III. Calculation of the A-C Resistance of a Pair Having Copper Conductors

The skin effect may be readily calculated by the use of the parameter x and curves of Fig. 24 which has been compiled from complicated exact formulas.[14]

$$x=2.845\sqrt{f/R_{dc}} \quad (29)$$

where f is the frequency in kc and R_{dc} the d-c resistance of the conductors in ohms per loop mile at the temperature under study.

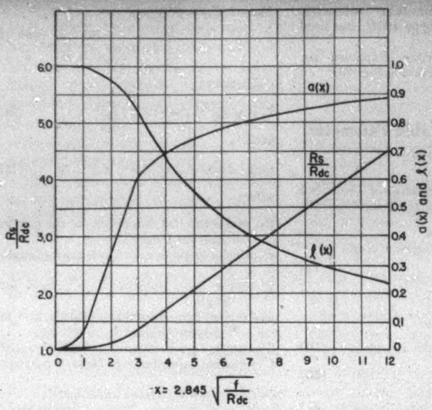


Fig. 24. Functions R_s/R_{dc}, $a(x)$ of equation 30 and $l(x)$ of equations 2 and 2(A) or 35 and 36

The d-c resistance may be calculated from the dimensions of the conductor using conventional formulas taking into account the temperature under study. One or two per cent should be added to this resistance which is an approximation to allow for the added length due to pairing and cabling. Having calculated x from equation 29, the resistance including skin effect is obtained by multiplying ratio R_s/R_{dc} of Fig. 24 for the corresponding x by R_{dc} and obtaining R_s, the a-c resistance including skin effect. The increment due to skin effect, ΔR_s, is equal to $R_{dc}(R_s/R_{dc}-1)$.

The a-c resistance including proximity effect as well as skin effect may be calculated by using the following formula due to Arnold.[15]

$$P=\sqrt{\frac{1}{1-A_0{}^2a(x)}}=\text{proximity factor} \qquad (30)$$

where

$A_0=d/s$
d = diameter of the conductor, inches
s = the center to center spacing of conductors, inches
$a(x)$ = see Fig. 24

(The 19 Awg 0.083, 19 Awg 0.066, and 22 Awg 0.083 cable tested during this program have insulated conductor diameters, or center to center spacing values, of 0.062 inch, 0.081 inch, and 0.045 inch, respectively.)

The approximate relationship, equation 30, has been derived by Arnold based on previous work done by Carson[16] and Dwight[17] and is simpler and can be evaluated in a shorter time than the more complicated and exact formulas developed by the others. Measurements indicate this formula is suitable for use in telephone cables and in the range up to 1,000 kc. Once having the factor P, the a-c resistance, including skin plus proximity effects, is calculated from the following formula:

$$R_{sp}=R_sP \qquad (31)$$

The increment due to proximity effect, ΔR_p, due to presence of the other conductor of the pair is equal to $R_{sp}-R_s$.

The component ΔR_t may be estimated by considering the surrounding pairs to be replaced by a copper tube and utilizing formulas developed by Meyerhoff[18] for losses in a sheath surrounding a pair of conductors. Meyerhoff has developed the following relationship for two wires in a sheath:

$$\Delta R_t=0.1913f'M_{21} \text{ microhm per foot}$$
$$=1.01\times10^{-3}f'M_{21} \text{ ohms per wire mile}$$

$$\Delta R_t=2.02\times10^{-3}f'M_{21} \text{ ohms per mile of a pair} \qquad (32)$$

where

$$M_{21}=4v\left[\frac{0.25}{1+v^2}+\frac{0.01563}{9+v^2}+\frac{0.00098}{25+v^2}+\ldots\right]$$

At frequencies in excess of 20 kc, M_{21} is equal to $1/v$ and ΔR_t may be expressed as follows:

$$\Delta R_t=\frac{2.02\times10^{-3}f'}{v}$$

ohms per mile of pair where

$$v=\frac{0.05p'f't}{\rho} \qquad (32A)$$

f' = frequency, cps
p' = mean diameter of the equivalent sheath, inches
t = thickness of the equivalent sheath, inches
ρ = resistivity of the material in the equivalent sheath, microhm-inches

If the frequency is sufficiently high, above 50 kc in this case, the eddy currents will flow near the surface of the wires and the well-known "depth of penetration" formula may be used to calculate t.

$$\delta(\text{inches})=t=\left[\frac{10\rho}{\mu f'}\right]^{1/2}$$
$$=3.162\sqrt{\rho/f'} \text{ for copper}$$

If this value of t is used and if p', the mean diameter of the equivalent sheath, is taken as $4s$, then ΔR_t may be reduced to the following formula:

$$\Delta R_t=0.0884\,A_0xR_{dc}$$

where

$$x=2.845\sqrt{f/R_{dc}} \text{ given in equation 24} \qquad (32B)$$

Equation 32(B) correlates well with data taken in the frequency range of about 100 kc and higher. At lower frequencies, the thickness t must be determined by equating the resistance of the wires surrounding the pair to the resistance of a solid tube, which for the case of copper, and if 12 or more wires are considered to be in parallel and represent the sheath, the thickness is simply $t=0.35/p'R_{dc}$ inches. However, rather than calculate t at the lower frequencies, ΔR_t has been modified to the following:

$$\Delta R_t=\frac{xA_0R_{dc}}{10} \quad \text{for } x>5 \qquad (33)$$

$$\Delta R_t=\frac{xA_0R_{dc}}{16} \quad \text{for } 2.5<x<5.0$$

$$\Delta R_t=\frac{xA_0R_{dc}}{24} \quad \text{for } x<2.5$$

where x and R_{dc} are defined in equation 29 and A_0 is d/s, as defined in equation 30. In the computed values given in this paper, ΔR_t was considered to be zero for values of x less than 1. The experimental check made to establish the reliability of these formulas was accomplished by measuring the a-c resistance of a pair at 80 F when contained in the complete cable and then removing the pair from the cable and again measuring the a-c resistance. The resistance of the pair removed from the cable deducted from the original resistance is ΔR_t. The data obtained on the three cables, each containing 100 pairs, are given in Table XVII.

The use of equation 33 should be reliable for folded corrugated aluminum sheath similar in resistance to the cable studied. Since the eddy current loss in these sheaths

Table XVII. Eddy Current Losses in Conductors and Sheath Surrounding a Pair (ΔR_i in Ohms per Mile) at 80 F

Frequency, Kc	19 Awg 0.066 Cable		19 Awg 0.083 Cable		22 Awg 0.083 Cable	
	Measured	Computed*	Measured	Computed*	Measured	Computed*
1	0	0.5	0	0.6	0	0.9
10	0.5–1.5	1.5	0.5–2.0	2.0	2–3	2.8
100	6–9	7.4	7–10	9.7	10–15	8.9
400	18–28	24	20–31	31	30–40	40
800	19–30	33	23–34	43	38–55	60
1,000	24–36	38	30–45	49	43–63	67

* Equation 33.

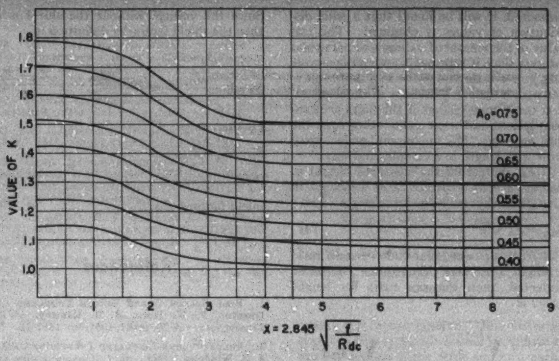


Fig. 25. Parameter K for equation 2(A) or 35 and 36

is so small, the equation should be reliable for cables with helically applied aluminum tapes which have a higher resistance. They have not been examined with cable having extruded aluminum sheaths; these latter cables may have different values of ΔR_t. Although this formula is based on cables having 100 pairs, it is expected to be reliable with a fair accuracy for any number of pairs, except perhaps for very small cables having fewer than 25 pairs.

Appendix IV. Inductance of a Pair Including Proximity Effect

The inductance of a single pair of wires in space including skin effect and proximity effect may be estimated from cable dimensions by the use of formulas developed by Carson[16] and Hebbert.[19] The following equation has been derived by Hebbert.

$$L=-4\log_e ks'+2KL_0 \qquad (34)$$

where

L = inductance of the pair per centimeter in centimeter-gram-second units

k = ratio $A_0/2$ (see equation 30)

$s'=\dfrac{1-\sqrt{1-4k^2}}{2k^2}$

K = correction factor of inductance L_0 for proximity effect and is a complex expression involving Bessel functions and is found from Fig. 25

L_0 = internal inductance of one wire with concentric return per centimeter in centimeter-gram-second units which includes skin effect only

L_0 has been defined by Hebbert[19] as follows:

$$L_0=\frac{2}{x}\,\frac{u_0u_0'+v_0v_0'}{u_1^2+v_1^2}$$

where x is given in equation 29, and u_0, u_0', v_0, v_0', u_1, and v_1 are the usual terms of the ber and bie functions related to the Bessel functions by the formula $u_n+iv_n=J_n(bi\cdot\sqrt{i})$ and $i=\sqrt{-1}$. Equation 34 may be rewritten as follows:

$$L=\left[1.485\log_{10}2s/d\times\frac{d^2}{2s(-\sqrt{s^2-d^2})}+0.161Kl(x)\right]\times 10^{-9} \qquad (35)$$

$$=\left[1.485\log_{10}\frac{d}{s-\sqrt{s^2-d^2}}+0.161\,Kl(x)\right]\times 10^{-3}\text{ henrys per loop mile of the pair} \qquad (36)$$

where

s = center to center spacing of the wires, inches

d = diameter of the wire, inches

$l(x)=2L_0$ and is obtained from Fig. 24

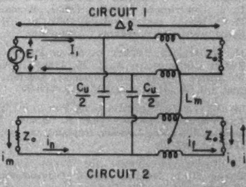


Fig. 26. Transverse crosstalk in a short element of paralleling pairs

Appendix V. Crosstalk Currents

By definition crosstalk is a ratio of power (P_2) induced in the disturbed circuit to the power (P_1) in the disturbing circuit. When expressed in db the crosstalk coupling loss is

$$10\log_{10}\frac{P_2}{P_1} \qquad (37)$$

When the characteristic impedances of both circuits are equal, the ratio of powers can be replaced by ratio of currents or voltages squared and the crosstalk loss becomes:

$$20\log_{10}\frac{E_2}{E_1}=20\log_{10}\frac{I_2}{I_1} \qquad (38)$$

Crosstalk can be expressed in crosstalk units, CU. Crosstalk unit is one millionth of one volt or ampere induced in the disturbed circuit by one volt or ampere, respectively, in the disturbing circuit. Crosstalk coupling loss in db, computed from CU, is

$$20\log_{10}\frac{E_2}{E_1}=20\log_{10}(\text{CU}\times 10^{-6})$$
$$=20\log_{10}\text{CU}-120 \qquad (39)$$

In 1907, Campbell[20,21] derived a convenient approximate expression for the crosstalk current or voltage ratio in an electrically short element of two long similar paralleling pairs of wires terminated in their characteristic impedance. This formula, while intended for use in studies of open wire, was soon adapted to cable studies. The derivation of the formula may be understood from Fig. 26.

The voltage E_1 drives a current $2i_e$ through the capacitance unbalance C_u and a current i_e appears at both ends of circuit 2. The current I_1 induces a voltage in circuit 2 by way of mutual inductance L_m and a crosstalk current i_m appears at both ends of circuit 2. The crosstalk currents i_e and i_m tend to add at the near end and to subtract at the far end of circuit 2.

The Campbell formula may be written:

$$\frac{I_2}{I_1}=\left[\frac{j\omega C_uZ_0}{8}\pm\frac{j\omega L_m}{2Z_0}\right] \qquad (40)$$

where

$\omega=2\pi f'$

f' = frequency, cps

C_u = capacitance unbalance in a short finite length Δl, farads

L_m = mutual inductance in a short finite length Δl, henrys

Z_0 = characteristic impedance of the line, ohms

The length Δl must be short enough to ignore phase change and long enough to involve approximately an integral number of times the least common multiple of the 2-pair twists. Then, with perfect twisting, the value of I_2/I_1 would be zero for any two pairs of unlike twist, of which one has an even number of twists in the length of common multiple.[22]

Much has been written[6,22] about the definition of C_u. Campbell's definition may

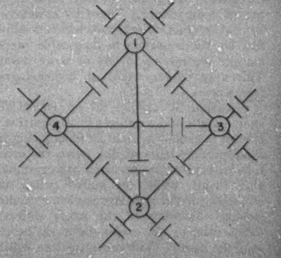


Fig. 27. Capacitance unbalance between two pairs

be understood by reference to Fig. 27 where *1–2* and 3–4 indicate two pairs: C_{13} indicates the direct capacitance between wires *1* and *3*, etc. Each of the wires *1* to *4* has many direct capacitances to other wires or to shield. C_u = capacitance unbalance $= C_{14} + C_{23} - C_{13} - C_{24}$.

This definition ignores certain second-order or "crosstalk of crosstalk" effects which, in well made cables, are unimportant, especially when no far-end balancing of crosstalk is employed.

In equation 40, each term contains $2\pi f'$ and tends to increase directly with frequency. However, the characteristic impedance Z_0 changes with frequency. The magnitude of Z_0 and of the angle θ both decrease with frequency as has been shown in Fig. 5.

For adjacent pairs C_u may have important values while for nonadjacent pairs C_u usually becomes negligible with the increase of number of separating pairs. At voice frequencies, the second, or I_m, term of equation 40 is usually neglected because of the high value of Z_0. This is particularly true if the pairs are loaded and have a higher impedance than Z_0.

At the carrier frequencies, Z_0 is relatively low. For adjacent pairs, the two terms of equation 40 become roughly equal and there is a marked tendency toward cancellation in the far-end crosstalk. For nonadjacent pairs, the second term becomes controlling and important. Nonadjacent pairs of like twist have been found to have serious far-end crosstalk at carrier frequencies and many different twist lengths are necessary to reduce mutual inductance.

For widely separated pairs of unlike twist, the simple Campbell formula may not apply. This covers "transverse" crosstalk, that is, the summation of crosstalk in a series of elemental lengths each Δl and neglects crosstalk between parts of primary and secondary circuits in different elemental lengths via tertiary circuits. This is known as "interaction" crosstalk[24] which may vary with length and frequency in a manner quite different from the simple transverse crosstalk within a section of cable between repeaters. This interaction crosstalk is not apt to be important since crosstalk currents arriving via transverse coupling are controlling except in cases where the transverse far-end crosstalk is substantially neutralized by crosstalk balancing units. This was done with the type K 12–60 kc carrier system. Crosstalk balancing has become obsolete because of new and cheaper methods of crosstalk reduction such as the use of compandors. There is, however, a type of interaction crosstalk which is amplified by the repeaters in long-cable carrier circuits repeated at one or more points. This may be important at higher carrier frequencies.

Equation 40 applies to an electrically short length and a process of summation is needed to establish crosstalk in long lengths. With parallel pairs of like twist, the crosstalk current ratio increases systematically with length while, for unlike twists, the crosstalk current ratio increases in a random fashion. In view of the many different twists used in the cables discussed herein, only the case of random summation will be discussed.

If, for a particular length, say 3,000 feet, a large number of values of current ratio are measured, it will be found that a wide distribution of values is obtained. The rms value of the measured values is of particular interest since it represents an average crosstalk power ratio and grows in a simple manner with length. The shape of the distribution curve may change if the pairs are random spliced and some value other than the rms, say the "1% maximum" value, may grow with length in a complicated way.

If F_Δ' represents the measured output to output far-end crosstalk ratio measured in a length Δ, the rms value for a length $n\Delta$ is

$$F_{n\Delta}' = F_\Delta'\sqrt{n} \tag{41}$$

If N_Δ' represents near-end crosstalk ratio in length Δ, so small that attenuation can be neglected, then the rms ratio for length $n\Delta = l$ is

$$N_l' = N_\Delta'\sqrt{\frac{1-e^{-4\alpha n\Delta}}{4\alpha\Delta}} = N_\Delta'\sqrt{\frac{1-e^{-4\alpha l}}{4\alpha\Delta}} \tag{42}$$

where α is attenuation in nepers per unit length and N_Δ' refers to the crosstalk ratio excluding the attenuation of crosstalk currents on their way to the near end. To include that attenuation, for instance when crosstalk current ratio N_{nl}' is computed from near-end crosstalk ratio N'_l measured on a length of cable l miles long, equation 42 is changed to:

$$N_{nl}' = N_\Delta'\sqrt{\frac{1-e^{-4\alpha nl}}{4\alpha\Delta}} = N_l'\frac{\sqrt{\dfrac{1-e^{-4\alpha nl}}{4\alpha\Delta}}}{\sqrt{\dfrac{1-e^{-4\alpha l}}{4\alpha\Delta}}}$$

$$= N_l'\sqrt{\frac{1-e^{-4\alpha nl}}{1-e^{-4\alpha l}}} \tag{43}$$

The near-end ratio grows more slowly than $\sqrt{n}$ because of the effect of attenuation. If n is very large, the numerator under the radical approaches unity.

Appendix VI. Shield Factor

The shield factor is an expression of the ratio of induced voltage in a shielded cable to that of the same cable unshielded. In the case of magnetic induction, where the shield and shielded conductors are closely associated in space, the external voltages induced longitudinally along the shield and along the shielded conductors are substantially equal. This voltage (E) is called the primary field. Assuming zero ground impedance, current i_s in the shield is equal to E/Z_{11} where Z_{11} is the ground return self-impedance of the shield. This current induces, along the shielded conductors, a counter voltage (E') called the secondary field which is equal to $i_s Z_{12}$ where Z_{12} is the ground return mutual impedance between the shield and the shielded conductors. The resultant voltage along the shielded conductors is then:

$$E - E' = E - i_s Z_{12} = E - \frac{E}{Z_{11}} Z_{12}$$

$$= E\left(1 - \frac{Z_{12}}{Z_{11}}\right)$$

$$= E\frac{(Z_{11} - Z_{12})}{Z_{11}} \tag{44}$$

Since the voltage without the shield is E, then the shield factor by definition is

$$S_f = \frac{E\dfrac{(Z_{11}-Z_{12})}{Z_{11}}}{E} = \frac{Z_{11}-Z_{12}}{Z_{11}} \tag{45}$$

Further, since

$Z_{11} - Z_{12} = Z_{ab}$

then

$$\text{shield factor} = Z_{ab}/Z_{11} \tag{46}$$

References

1. **Bell System Cable Sheath Problems and Designs, F. W. Horn, R. B. Ramsey.** *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1811–16.

2. **Polyethylene-Insulated Telephone Cable, A. S. Windeler.** *Ibid.*, pt. I (*Communication and Electronics*), vol. 73, May 1954, pp. 106–11.

3. **Inductance Calculations: Working Formulas and Tables (book), Frederick W. Grover.** D. Van Nostrand Company, Inc., Princeton, N. J., 1946, p. 274.

4. **Interaxial Spacing and Dielectric Constant of Pairs in Multipaired Cables, J. T. Maupin.** *Bell System Technical Journal*, New York, N. Y., vol. 30, July 1951, pp. 652–67.

5. **The Control of Noise and Crosstalk on N1 Carrier Systems, A. J. Aikens, C. S. Thaeler.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 72, Nov. 1953, pp. 605–11.

6. **Crosstalk on Carrier Cables, K. E. Latimer.** Wire and Cable Symposium, Asbury Park, N. J., 1955.

7. **Possibilities of Carrier Working on De-Loaded Voice-Frequency Cables: An Analysis of Cable Properties, H. Feiner, L. J. E. Kolk.** *Communication News*, London, England, vol. 16, Feb. 1956, pp. 43–59.

8. **Cable Crosstalk—Effect of Non-Uniform Current Distribution in the Wires, R. N. Hunter, R. P. Booth.** *Bell System Technical Journal*, vol. 14, Apr. 1935, pp. 179–94.

9. **The Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shields, S. A. Schelkunoff.** *Ibid.*, vol. 13, Oct. 1934, pp. 532–79.

10. **Shielding of Communication Cables, F. H. Gooding, H. B. Slade.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, July 1955, pp. 378–87.

11. **Wave Propagation in Overhead Wires with Earth Return, ohn B. Carson.** *Bell System Technical Journal*, vol. 5, Oct. 1926, pp. 539–54.

12. **Communication Engineering (book),. William Littell Everitt.** McGraw-Hill Book Company, Inc., New York, N. Y,. second edition, 1937, pp. 168–71.

13. **High Frequency Alternating Currents (book), Knox McIlwain, J. G. Brainerd.** John Wiley & Sons, Inc., New York, N. Y., second edition, 1939, p. 350.

14. **Radio Instruments and Measurements (book), U. S. National Bureau of Standards.** U. S. Government Printing Office, Washington, D. C., second revised edition 1937, p. 309 (National Bureau of Standards Circular *74*).

15. **Proximity Effect in Solid and Hollow Round Conductors, A. H. M. Arnold.** *Journal*, Institution of Electrical Engineers, London, England, vol. 88, pt. II, Aug. 1941, pp. 349–59.

16. **Wave Propagation over Parallel Wires: The Proximity Effect, John B. Carson.** *Philosophical Magazine*, London, England, vol. 41, Apr. 1921, pp. 607–33.

17. **Proximity Effect in Wires and Thin Tubes, H. B. Dwight.** *AIEE Transactions*, vol. 42, Sept. 1923, pp. 961–70.

18. **Pipe Losses in Nonmagnetic Pipe, L. Meyerhoff.** *Ibid.*, pt. III (*Power Apparatus and Systems*), vol. 72, Dec. 1953, pp. 1260–75.

19. THE TRANSMISSION CHARACTERISTICS OF TOLL TELEPHONE CABLES AT CARRIER FREQUENCIES, C. M. Hebbert. *Bell System Technical Journal*, vol. 20, July 1941, pp. 293–330.

20. DIRECT CAPACITY MEASUREMENTS, G. A. Campbell. *Ibid.*, vol. 1, Nov. 1922, pp. 18–38.

21. DR. CAMPBELL'S MEMORANDA OF 1907 AND 1912. *Ibid.*, vol. 14, Oct. 1935, pp. 558–72.

22. ÜBER NEBENSPRECHSTÖRUNGEN IN FERNSPRECHKABELN, H. Schiller. *Elektrische Nachrichten-Technik*, Berlin, Germany, vol. 8, 1931, pp. 114–21; vol. 9, 1932, pp. 81–91.

23. CROSSTALK, H. C. S. Hayes, R. A. Seymour, P. R. Bray. The Institution of Post Office Electrical Engineers, London, England, 1938, 48 pp.

24. OPEN-WIRE CROSSTALK, A. G. Chapman. *Bell System Technical Journal*, vol. 13, Jan.-Apr. 1934, pp. 19–37, 195–238.

◆

Discussion

J. V. Buscemi (Rural Electrification Administration, U. S. Department of Agriculture, Washington, D. C.): Crosstalk characteristics have been presented in this paper in sufficient detail regarding frequency and position of pairs in the cable that the telephone system application engineer has an adequate guide as to which of the pair combinations may contribute controlling crosstalk. Under certain conditions, at repeater points, however, not transverse but interaction crosstalk may be controlling. This may occur when the repeater gain desired is higher than the conditions of interaction crosstalk allow.

It would be of value if the authors could supply data on the magnitude of repeater gain at which interaction crosstalk becomes equal to or more pronounced than the far- and near-end transverse crosstalk indicated in the paper.

G. S. Eager, Jr., L. Jachimowicz, I. Kolodny, and D. E. Robinson: We feel that Mr. Buscemi's point is very well taken since interaction crosstalk (coupling loss) may determine the permissible repeater spacing and gain. In answer to this question, the following comments and data are furnished.

In effect, near-end, near-end interaction crosstalk is near-end crosstalk repeated twice and amplified at the repeater point. The signal amplified at the repeater point and sent in one direction, for instance, east, appears as a near-end crosstalk in a pair or pairs of the local circuits which are continuous at the repeater point. The disturbing signal of this crosstalk travels back into the western section of the cable and by a path of near-end crosstalk appears on the western termination of the carrier pair in which the signal arriving for amplification is weak. Interaction crosstalk will not achieve the controlling magnitude of its crosstalk coupling loss expressed in db minus the value of repeater gain in db but will be equal or larger than controlling transverse far-end crosstalk coupling loss of the repeater section.

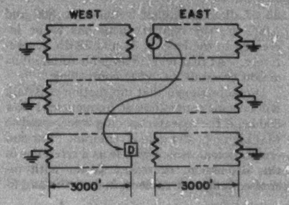


Fig. 28. Schematic diagram of circuit used in measuring near-end, near-end interaction crosstalk

The interaction crosstalk coupling loss given in Table XVIII was measured on the same two 100-pair 0.083 μf per mile 19 Awg cables, each 3,000 feet long, used for the other crosstalk measurements described in this paper. The cables were spliced straight color to color, as shown schematically in Fig. 28. Pairs not under direct test were terminated at both ends of the 6,000-foot length in resistors, center taps of which were grounded, so as to simulate a long line and suppress reflection from termination of phantoms and other tertiary circuits. The disturbed and the disturbing pairs were cut in the middle of the 6,000 foot length, and terminated into precision networks. The disturbing signal was fed into the eastern 3,000-foot section and the signal received in the disturbed pair was measured, arriving at the 3,000-foot mark of the western section of the cable. The ends of inactive halves of the disturbed and disturbing pairs at the middle of the cable were terminated in resistors, the center taps of which were not grounded.

It can be noted that the test was conducted, on purpose, in most unfavorable conditions with all pairs, except for two under test, uninterrupted at the repeater point. No clear pattern of increase in crosstalk coupling loss, with increase of separation between pairs, has been observed. A disturbing pair induces crosstalk in a multitude of tertiary circuits in the eastern portion of the cable which, in turn, induces voltages in pairs of the western portion of the cable, without much difference whether pairs are closely associated in space in the eastern or in the western section of the cable.

Table XIX. Measured Rms Crosstalk Losses in Db Corrected to 10,000-Foot Adjacent Pairs

Frequency, Kc	Interaction Crosstalk Coupling Loss: Before Amplification	Interaction Crosstalk Coupling Loss: Equal Level after Amplification	Far-End Equal Level Crosstalk Loss	Difference
		Adjacent Pairs		
100	104	93	69	−24
300	96	77	60	−17
500	85	59	55	−4
800	79	46	50	+4
1,000	75	37	49	+12
		Alternate Pairs		
100	104	93	67	−26
300	97	78	65	−13
500	90	64	61	−3
800	84	51	56	+5
1,000	79	41	55	+14
		Two Separators		
100	104	93	77	−16
300	98	79	67	−12
500	91	65	62	−3
800	85	52	59	+7
1,000	81	43	57	+14
		Four Separators		
100	109	98	87	−11
300	98	79	79	0
500	92	66	74	+8
800	87	54	68	+14
1,000	82	44	65	+21
		Center to First Layer		
100	102	91	71	−20
300	92	73	62	−11
500	84	58	55	−3
800	85	52	53	+1
1,000	83	45	52	+7

An evaluation of the significance of interaction crosstalk for a line with 10,000-foot spacing of repeaters is given in Table XIX. The far-end crosstalk for such sections will be as given in Tables V and VI:

$$15-10\log_{10}\frac{10,000}{30,000}$$

Near-end, near-end interaction crosstalk

Table XVIII. Rms Measured Interaction Crosstalk Coupling Loss, in Db

Frequency, Kc	Pair Combinations: Adjacent Pairs: Center	Adjacent Pairs: 1st Layer	Adjacent Pairs: 2nd Layer	Adjacent Pairs: Grand Rms	Alternate Pairs: 1st Layer	Alternate Pairs: 2nd Layer	Alternate Pairs: Grand Rms	Two Separators: 1st Layer	Two Separators: 2nd Layer	Two Separators: Grand Rms	Four Separators, 1st and 2nd Layers	Center to 1st Layer
10	123	113	130	118	118	123	121	119	124	122	122	118
30	108	98	113	103	103	107	105	105	110	108	108	101
60	107	103	111	106	106	102	103	108	111	109	108	103
100	106	103	108	106	105	107	106	104	108	106	111	104
150	119	97	103	100	98	106	101	96	107	99	103	101
300	99	95	99	97	98	98	98	97	101	99	99	93
500	78	94	88	86	92	91	91	94	91	92	93	85
800	78	81	81	80	83	89	85	84	87	86	88	86
1,000	73	76	78	76	77	82	80	80	84	82	83	84

loss will decrease much less with the increased length of cable, the decrement being twice the small decrement of near-end crosstalk loss. As shown in Fig. 20, the value of near-end crosstalk loss changes little from the value measured on a 3,000-foot length at frequencies above 100 kc.

It can be seen from Table XIX that in 10,000 feet of 19 Awg 0.083 cable near end, near-end interaction crosstalk is not controlling as long as line attenuation losses and, therefore, repeater gains do not exceed some 25–27 db. At higher frequencies, 800 and 1,000 kc, the repeater gain has to be increased to the level of 33 and 38 db, respectively. At that level of amplification, interaction crosstalk becomes controlling.

It can be shown that interaction crosstalk will become controlling at frequencies below 800 kc, if the repeater spacing is increased so that high repeater gain is required. Far-end crosstalk loss, which we use as a reference in this discussion, decreases slower with increased length of cable than does the amplified interaction crosstalk loss. When the length of cable is increased n times, the far-end crosstalk loss, expressed in db, will be $F_n = F_0 - 10 \log_{10} n$ and amplified interaction crosstalk $N_{tn} = N_{t0} - \alpha n$ where N_{t0} is practically constant for longer lengths of cable and α is line attenuation of the $1/n$ length of cable. As n is increased, there will always be a point at which $N_{tn} < F_n$, that is, repeater spacing at which near-end, near-end interaction crosstalk will become controlling.

An Electric Analog of Heat Flow in Power Transistors

JAMES REESE
NONMEMBER AIEE

W. W. GRANNEMANN
NONMEMBER AIEE

J. R. DURANT
NONMEMBER AIEE

Synopsis: An electric analog of heat flow in alloyed-junction power transistors was constructed to facilitate a study of instantaneous crystal temperatures under various pulsed conditions. An equivalent T-section transmission line using lumped constants was chosen as a suitable analog. The derivation of the transmission-line analog is given, along with construction and operating procedures. A short comparison of analog and experimental data is included.

IN THE DESIGN of circuits utilizing power transistors, the maximum allowable instantaneous junction temperature is often the major limitation on the power carrying capabilities of the circuit. Transistor failure may result if the junction temperature exceeds, even momentarily, a certain maximum value.

While the maximum safe-operating junction temperature is usually given by the manufacturer, it is often difficult to determine what the junction temperature actually is under various excitation conditions. An electric analog of heat flow in power transistors can be designed to give, immediately, the temperatures throughout the crystal under a wide variety of excitation conditions. Further, the analog may be used to study ways of extending the maximum safe-operating temperature.

Fourier's heat-flow equation in one-dimension is analogous to the equation for voltage variation along a transmission line. By reducing the heat-flow model of a power transistor to a one-dimensional model, one may simulate the transistor with a transmission line. To simplify construction and operation, the continuous transmission line may also be simulated with an equivalent T-section lumped-parameter line.

Derivation of Analog

A complete analysis of the heat flow in an alloyed-junction power transistor would be extremely complex. Several simplifying assumptions are possible, however, which do not greatly affect the accuracy of the results.

The first of these assumptions, that radiation losses are negligible, can be validated by use of

$$q = A\sigma E(T^4 - T_0^4) \qquad (1)$$

where

q = total heat loss due to radiation, erg/sec (second)
σ = Stefan-Boltzmann constant, 5.672 (10^{-5}) erg/sec centimeter^2 degrees Kelvin^4
E = emissivity of radiator crystal
T_0 = temperature of receiver (case), degrees Kelvin
T = temperature of radiator, degrees Kelvin
A = total area of radiator, centimeter^2

The crystal is assumed to be a perfect radiator with an emissivity of one. The temperature difference between the crystal and the case is assumed to be sufficient to damage the crystal (approximately 80 C (degree centigrade) for most power transistors). Solving equation 1 for q with these assumptions gives the maximum possible heat loss caused by radiation.

The maximum heat loss due to radiation for the Minneapolis-Honeywell type *H-6* transistor is approximately 0.082 watt for a total average power input of approximately 40 watts. This 40 watts of input power is sufficient to damage the transistor regardless of the heat sink used. Maximum radiation losses, then, will amount to only 0.205% of the total power dissipated for the *H-6* transistor.

Since heat loss due to natural convection is approximately proportional to the 5/4 power of the temperature difference,[1] it will certainly be less than the loss due to radiation. Doubling the radiation loss, then, will give a value for radiation and natural convection losses which will be higher than is actually the case. Thus the total heat losses due to radiation and natural convection are less than 0.41% of the total power dissipated for the *H-6* transistor.

The heat loss caused by conduction through the gas surrounding the crystal is easily shown to be very small by use of

$$q = \frac{KA\Delta T}{d} \qquad (2)$$

where

q = heat loss, calorie/second
K = heat conductivity of the gas, calorie/second-C-centimeter
A = total area of the crystal, centimeter^2
ΔT = temperature difference between crystal and case, C
d = distance from crystal to case, cm (take d a minimum to obtain q maximum)

The heat conducted away through the gas in the *H-6* transistor is approximately 0.02 watts maximum for a total average power input of 40 watts.

Maximum heat loss due to radiation, natural convection, and conduction through the gas, then, will be less than 1/2 of 1% of the total power dissipated for the *H-6* transistor. These losses should not

---

Paper 59-766, recommended by the AIEE Solid State Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted October 14, 1958; made available for printing April 1, 1959.

JAMES REESE, W. W. GRANNEMANN, and J. R. DURANT are with the University of New Mexico, Albuquerque, N. Mex.

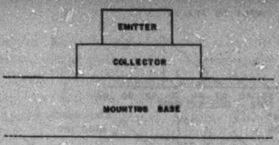


Fig. 1. The thermal model

be over 1% of the total power dissipated for any gas-filled power transistor.

By neglecting these heat losses, the heat flow problem is greatly simplified. If it is assumed that all elements of the crystal are homogeneous and cylindrical in shape the one-dimensional thermal model shown in Fig. 1 may be used.

Fourier's equation of the heat flow in this one-dimensional model may be expressed as

$$\frac{\partial T}{\partial t}=\alpha_T\frac{\partial^2 T}{\partial x^2} \tag{3}$$

where

T = temperature
X = distance from junction
α_1 = thermal constant of crystal
t = time

Equation 3 is the standard equation for voltage variation along a transmission line

$$\frac{\partial V}{\partial t}=\alpha_e\frac{\partial^2 V}{\partial x^2} \tag{4}$$

with voltage replaced by temperature and the electricl constant α_e replaced by the thermal constant α_T.

A transmission line, then, is one analog for the one-dimensional heat flow problem. An equivalent T-section transmission line using lumped constants was chosen as the easiest to construct and operate.

Table I. Useful Analogies in Selection of Component Sizes and Number of Sections in Constructing the Analog

	Electrical	Thermal
Conservation of Scalar quantity	charge, coulombs	heat, calories
Scalar point function	electric potential, volts	temperature, degrees
	Ohm's law	Fourier's law
Resistance concept	$R_e=dV/I$	$R_t=dT/q$
	R_e = electric resistance	R_t = thermal resistance
	I = current through R_e	q = heat flow through R_t
	dV = difference in potential across R_e	dT = difference in temperature across R_t
Capacity concept	$C_e=Q_e/dV$	$C_t=Q_t/dT$
	C_e = electric capacity	C_t = thermal capacity
	Q_e = charge stored in C_e	Q_t = heat stored in C_t
	dV = rise in electric potential of C_e because of Q_e	dT = rise in temperature of C_t due to Q_t
One-dimensional form of heat conduction equation for solids	t_e = time in electric circuit	t_t = time in thermal circuit
	x_e = distance along cable	x_t = distance along flow path
	V = electric potential at X_e at time t_e	T = temperature at x_t at time t_t
Flow	amperes, coulombs/second	calorie/second
Capacity	farads, coulombs/volt	calorie/C
Resistance	ohms, volts/coulomb/second	C/calorie/second

Construction of Analog

The question of component size and the number of sections required to construct our analog immediately arises. The analogies given in Table I are useful in answering these questions.

Using the procedure outlined in Creamer,[2] it was determined that 50 identical T sections for both the emitter and collector and 10 T sections for the base region would simulate the transmission line up to a frequency of 50 kc. This frequency range was extended somewhat by tapering the line. In constructing the analog, considerable choice is allowed in the selection of component sizes if the proper scaling technique is employed. The following scaling constants may be employed:

L = scaling constant for voltage and temperature
N = scaling constant for time
$1/M$ = scaling factor for capacitance
NM = scaling constant for resistance
L/NM = scaling factor for current and heat flow

Since the analog must be fed by a current source, (with no input to analog, the source must not provide a discharge path) the factor NM is of primary importance in the design of the feed source. The output impedance of the feed source must be several (10 or more) times the impedance of the analog. Taking the output impedance (preferably a pure resistance) of an easily realizable current generator, we fix the value of NM with inequality

$$R_e(NM)\leq\frac{Z_0}{10}$$

using this value of NM, we calculate R_e with the equation

$$R_e=\frac{R_t}{NM}$$

where R_t is the total thermal resistance of either the collector or emitter.

R_t is given by

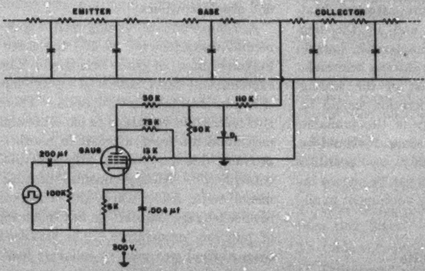


Fig. 2. Power transistor heat flow analogy and current generator

Collector = 50 T sections
Emitter = 50 T sections
Base = 10 T sections

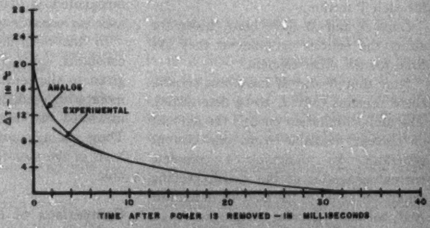


Fig. 3. Measurements of versus "off" time

Pulse width = 10 milliseconds
Period = 50 milliseconds
Peak power = 42 watts

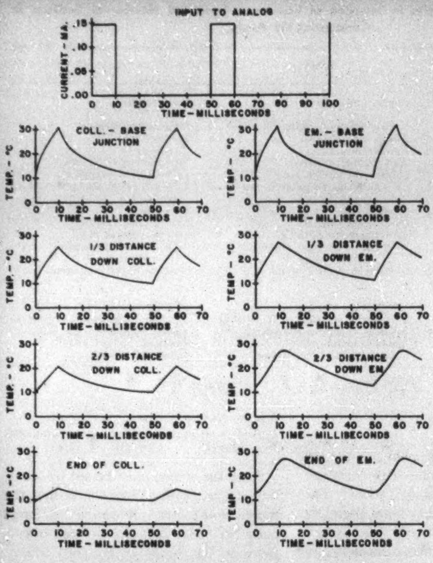


Fig. 4. Analog waveforms for rectangular pulsed input

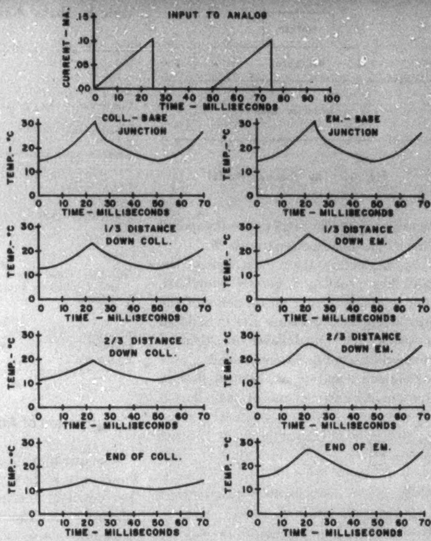


Fig. 5. Analog waveforms for sawtooth input

$$R_t=\frac{d}{KA}$$

where

d = depth
A = area
K = heat conductivity

This value of R_t is divided by twice the number of T section to obtain the resistance for each arm of each T section.

Since $C_e=MC_t$, one is free to choose a value of m which will give a convenient value for the electric capacity, when the factor NM is kept constant. The value so chosen is divided by the number of T sections to obtain the value of capacity for each T section.

Once N and M have been chosen for either the emitter or collector they are fixed for the other element.

Now that N and M have been chosen, there remains only L to be determined. The major limitation on L is the amount of current available from our current generator. By designing a complete current generator at this point, keeping mind the value of Z_0 used to determine NM, solve for L using

$$L=\frac{NM}{I_{\max}}q_{\max}$$

where $I_{\max}$ is the maximum current available from our current generator and $q_{\max}$ is the maximum heat flow of interest. LV now corresponds to the heat flow into the crystal necessary to produce this temperature. Thus a variation in the current from the generator corresponds to a variation in heat flow into the crystal and a variation of voltage along the line corresponds to a variation of temperature through the crystal.

The scaling constants used to determine the electrical constants of the T sections for the base, collector, and emitter are also used to determine the electrical constants needed to simulate the copper mounting base and any attached heat sinks. The collector end of the line terminated in its characteristic impedance corresponds to an infinite heat sink.

In the determination of the scaling constants, some consideration should be given to the sensitivity of the available monitoring equipment since V should be measured as accurately as possible. These scaling constants can be chosen on basis of the monitoring equipment available.

Comparison of Results

Rather than compare exact experimental and analog temperatures, it is simpler to compare values of ΔT. ΔT is defined as the difference between maximum junction temperature (the temperature at the end of the power pulse) and minimum junction temperature (the temperature at the beginning of the power pulse). Using ΔT for comparison purposes allows the clamping diode, D_1 in Fig. 2, to clamp the input pulse to a zero reference level. This clamping removes the ambient temperature factor from the analog, thus simplifying the analog operating procedure. Further, we are able to utilize the maximum sensitivity of our monitorying equipment, since we are interested only in voltage differences and not absolute values.

A comparison of analog and experimental measurements of ΔT, for a rectangular pulse, is given in Fig. 3. The experimental curve shown is an average of measurements for several units. There is considerable variation in experimental measurements from one unit to another, probably due to nonuniformity of crystal construction. All experimental measurements were made using the collector reverse-leakage current as the indicator of junction temperature. The standard oven method was used to calibrate leakage current versus junction temperature.

As shown in Fig. 3, there is good agreement in measurements of over most of the cooling cycle. However, no reliable experimental data are available for the period immediately following the removal of the power pulse. Considerable difficulty has been encountered in attempting to separate, experimentally, the electrical and thermal effects during this period. It is expected that, once experimental data is available, analog and experimental measurements will essentially agree throughout the cooling cycle.

The difficulty in obtaining experimental data immediately following the power pulse is eliminated with the analog. Measurements of temperature throughout the cycle, both heating and cooling portions, are readily available using the analog.

The complete power cycle, as given by the *H-6* analog for two different waveforms, is shown in Figs. 4 and 5. The variations in temperature, with time, are shown for various points in the crystal. Note that the emitter and collector junctions are at essentially the same temperature throughout the cycle. This indicates that if a transistor were operating near the thermal failure temperature, any increase in power applied to the emitter junction might cause the emitter junction to fail since its temperature could be higher than that of the collector.

Several transistors were damaged under collector pulsed conditions. The power was removed before thermal runaway was completed. These transistors showed damage to the emitter junction only. Usual failures under thermal runaway conditions result in damage to both junctions.

Thermal runaway conditions may by simulated with the analog by adding a second current generator to the input. The second current generator would correspond to the power input to the crystal caused by leakage current.

Some of the other advantages of the analog as compared to experimentally obtained data are:

1. Measurements of temperatures anywhere in the crystal are possible with the analog. The only temperatures immediately available experimentally are the temperature at the junction and the temperature of the mounting base.
2. Changes in the physical geometry (size of base fin, collector termination, size of emitter and collector, etc.) are easily made on the analog.
3. Ambient temperature effects may be noted on the analog without ovens or refrigeration.
4. Changes in transistor characteristics with age and environment which necessitate constant calibration for experimental data are eliminated in the analog.
5. Measurements of power input necessary to raise junction temperature to a dangerous level under various pulsed conditions can be made, with the analog, with no risk of damaging transistors.
6. With the analog, the time required to obtain specific pulse data is greatly reduced.
7. The effect of various cooling techniques may be investigated, quite easily, with the analog.
8. In using the analog, a mathematical representation of the waveform under study is not necessary. If the waveform can be generated and applied to the analog the temperature may be determined at any point in the crystal.

References

1. Conduction of Heat in Solids, H. S. Carslaw, J. C. Jaeger. Oxford University Press, New York, N. Y., 1947, p. 16.

2. Communication Networks and Lines, Walter J. Creamer. Harper & Brothers, New York, N. Y., 1951, pp. 208–12.

Transistor-Magnetic Control Circuits for Aircraft Electric Systems

A. W. PRATT
ASSOCIATE MEMBER AIEE

THE STRINGENT requirements for military aircraft electric systems necessitates the optimum utilization of the latest component developments. The circuitry described in this paper is intended to combine semiconductor diodes and transistors with square-loop magnetic cores in a manner that results in the maximum use of their desirable characteristics. The memory capability of square-loop magnetic cores and the switching characteristics of transistors are employed in a complementary manner. As a result, the following advantages are obtained:

1. The transistor-magnetic high-phase and low-phase sensing circuit inherently senses the "true average" value of the system alternating voltages. In other words, it senses the total amount of volt-seconds underneath the positive half cycles of the voltages. This is in contrast to conventional circuits which sense either the peak value or only a small fraction of the volt-seconds underneath the waveform.
2. Because only the switching characteristics of the transistors are utilized, circuit operation is practically independent of variations in transistor characteristics.
3. The circuits use only small, static, reliable, and rugged components. Capacitors, which are used in conventional circuits, are replaced in this circuit by square-loop magnetic cores which are less temperature sensitive and much more reliable.

The transistor-magnetic circuits that are introduced in this paper are

1. High- and low-phase sensing circuit.
2. Overvoltage protection circuit with inverse time delay.
3. Undervoltage protection circuit.
4. Unbalanced phase protection circuit.
5. Fixed time-delay circuit.

The block diagram of Fig. 1 shows one possible combination of these circuits to provide protection for a 3-phase system. The pulse delay reactor shown simply delays any output pulse from the undervoltage circuit that may occur so that it will not be applied to the fixed time-delay circuit when this circuit is being reset and is consequently unable to respond. The circuitry for accomplishing this function will not be discussed in detail since it may assume a number of conventional forms. The latching circuit shown may also be conventional and so it will not be described.

High- and Low-Phase Sensing Circuit

In order to provide maximum protection in a multiple-phase electric system it is desirable to monitor the highest and the lowest phase voltages of the system. It is not sufficient to monitor the average value of the phase voltages by means of a 3-phase bridge rectifier for

---

Paper 59-765, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 6, 1959; made available for printing, April 2, 1959.

A. W. Pratt, formerly with Jack & Heintz, Inc., Cleveland, Ohio, is now with Leland Airborne Products, Vandalia, Ohio.

The author wishes to acknowledge the generous assistance of Mr. R. J. Wurm of Jack & Heintz, Inc. in obtaining test data on these circuits.

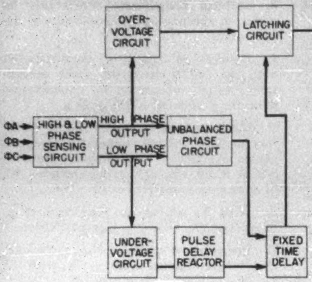


Fig. 1 (left). Block diagram of one combination of transistor-magnetic control circuits

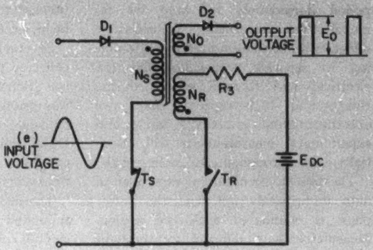


Fig. 2 (right). Simplified, single-phase pulse-forming circuit

a 3-phase system for example. Fault conditions are possible which will cause one phase voltage to be considerably higher than the others, while causing the average value of the phase voltages to be less than the specified maximum value It is obvious that this condition may produce harmful effects on the system load if the overvoltage protection circuit senses the average value of the 3-phase voltages. However, if the circuit sensed only the highest of the phase voltages, it is apparent that none of the phase voltages would exceed the specified maximum value. Consequently, every portion of the load would be protected against overvoltage.

A parallel example could be cited for an undervoltage condition demonstrating the desirability of having low-phase sensing for the undervoltage protection circuit.

Another requirement in some a-c systems is that the average value of each phase voltage be monitored rather than the peak value. One reason for this preference is that the average value of the voltage indicates more accurately the heating effect of the voltage upon load devices.

The high- and low-phase sensing circuit that will be described senses the true average values of the highest and the lowest phase voltages of a 3-phase system. By "true average" sensing of an alternating voltage is meant that the circuit can sense essentially all of the volt-seconds underneath the waveform of the voltage. This is in contrast to the sensing of only a small portion of the volt-seconds underneath the waveform that is performed by some circuits that are termed average sensing

SINGLE-PHASE PULSE-FORMING CIRCUIT

The circuit of Fig. 2 is capable of producing a rectangular voltage pulse at its output terminals which contains an amount of volt-seconds that is proportional to the volt-seconds underneath the positive portion of the input voltage waveform. In order to do this, it is necessary that the switches T_s and T_R, which are intended to simulate transistors, be operated in a specified manner.

For purposes of explanation, assume that the input voltage is approximately a sine wave. During the sensing period switch T_s is closed and T_R is open. Because of diode D_1 only the positive portion of the input voltage is applied to winding N_s. Therefore the flux change produced on the core is proportional to the amount of volt-seconds underneath the positive half-cycle of the input voltage. During this time diode D_2 blocks current flow in winding N_o and the open switch prevents current flow in winding N_R. Consequently, there is no induced current flowing in the input circuit during the sensing period. The resistive voltage drop in this circuit is therefore caused mainly by the magnetizing current required by N_s and is negligible.

Sometime during the negative half-cycle of the input voltage, switch T_s is opened and T_R is closed. The voltage E_{dc} is then applied to winding N_R until the core saturates in the negative direction. It should be noted at this point that resistor R_3 is small so that the voltage drop produced across it by the magnetizing current required by N_R prior to saturation is negligible in relation to E_{dc}. Prior to saturation of the core a voltage is induced in winding N_0, with the polarity shown, that is proportional to E_{dc}. Since the change in flux during this reset period must be equal to the change in flux that occurred during the previous sensing period, the voltage pulse that occurs at the output terminals contains an amount of volt-seconds that is proportional to the volt-seconds underneath the positive half cycle of the input voltage. It should be noted that the open switch T_s prevents current flow in winding N_s during this reset period.

THREE-PHASE PULSE-FORMING CIRCUIT

In order to produce the required pulse for each of the three phases the circuit shown in Fig. 3 is used. It is noted that the input winding of the transformer for phase B is inverted with respect to the similar windings of phases A and C. It will be shown later why this is necessary.

The circuit operates in the following manner. As described for the single-phase circuit, each of the three pulse-

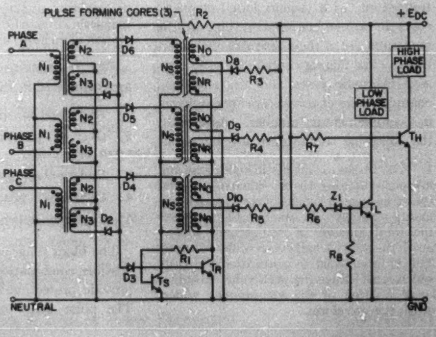


Fig. 3. High-phase and low-phase sensing circuit

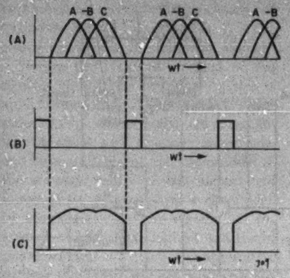


Fig. 4. Waveform sketches showing switching sequence of transistors in high- and low-phase sensing circuit

A—Rectified input voltages
B—Base voltage on reset transistor
C—Base voltage on sensing transistor

forming cores has a flux change during the sensing period that is proportional to the volt-seconds underneath the positive half-cycle of its phase voltage. This statement is made in reference to the secondary voltages of the three input transformers. The phase relationship of these voltages is shown in Fig. 4. Because the voltage of phase *B* is inverted, as described previously, the phase difference between the three voltages is 60 degrees. Transistor T_s remains closed until each of the three positive half-cycle voltages is "read" by its respective pulse-forming core.

The three cores are reset in parallel during the 60-degree period when all of the input transformer secondary voltages are negative as shown in Fig. 4.

During this "readout" period each of the three output windings produces a pulse that has a volt second content that is proportional to the volt-second content of its phase voltage. It is shown in Appendix I that this pulse is also proportional to the "true average" value of the phase voltage and inversely proportional to the system frequency.

Method of Sensing High- and Low-Phase Pulses

The high-phase output signal is obtained by applying the sum of the three output voltages to the base of the high-phase transistor in such a manner that any one of the output voltages by itself is sufficient to saturate that transistor. This is accomplished by connecting the output windings of the three cores in series so that their voltages add and setting the value of the base resistor R_7 so that the high-phase transistor is saturated when any one of the output voltages is present. It is also necessary that the current flowing in the output windings be less than the magnetizing current required by these windings so there is no positive change in flux of the cores during the reset period.

As a result of the foregoing considerations, the length of time that the high-phase transistor is on is seen to be equal to the width of the high-phase pulse. If the high-phase load is placed in series with the collector of this transistor, and the voltage E_{dc}, which is the same as that voltage which was used to reset the cores, is applied to this load, the volt-seconds of the pulse which occurs across the load during the reset period is proportional to the average value of the high-phase input voltage and inversely to the system frequency.

In order to sense the low-phase pulse, the sum of the three output voltages is applied through a Zener diode Z_1 to the base of the low-phase transistor. The breakdown voltage of Z_1 is such that the base feed of the low-phase transistor is insufficient to allow it to conduct when any one of the three output voltages disappears. If the load for this transistor is placed in series with its collector and the voltage source E_{dc} as for the high-phase transistor, the volt-second content of the voltage pulse produced across the low-phase load will be proportional to the average value of the low-phase system voltage and inversely proportional to the system frequency.

Method of Producing Proper Switching of Transistors

As described in the foregoing it is required that transistor T_R conduct during the 60-degree period when the voltages of phases *A* and *C* are negative. The "and" circuit, which consists of diodes D_1, D_2, D_3 and resistor R_3, applies a voltage to the base of T_R only for this condition. This voltage biases transistor T_R to saturation during the proper period.

Resistor R_1 causes the sensing transistor T_s to be biased to saturation when transistor T_R is off, and causes it to be biased off when transistor T_R is biased to saturation. Thus, it is seen that the circuit of Fig. 3 provides for proper switching of the transistors.

Overvoltage Circuit with Inverse Time Delay

In this circuit the high-phase pulse that was described as the high-phase output pulse for the high- and low-phase sensing circuit is compared to a reference pulse whose width is equal to one half of the period of the sensed voltage and whose magnitude is regulated at a constant voltage. When the average value of the highest phase voltage exceeds a specified value, an output signal is produced after a time delay which is inversely proportional to the difference between the average value of the highest phase voltage and the magnitude of the voltage of the reference pulse. A detailed analysis of the operation of this circuit is given in Appendix II. A general description of operation is given in the following.

The circuit is shown in Fig. 5. The

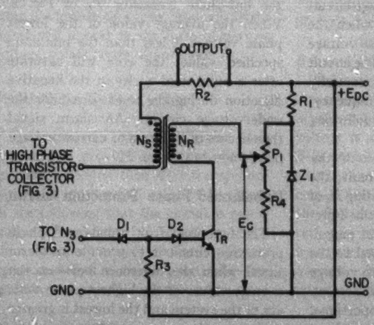


Fig. 5 (left). Overvoltage circuit with inverse time delay

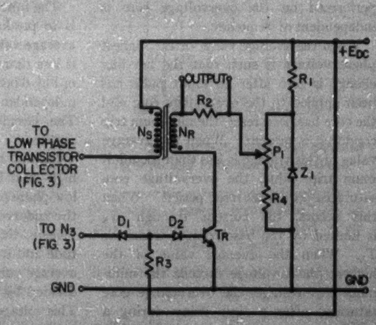
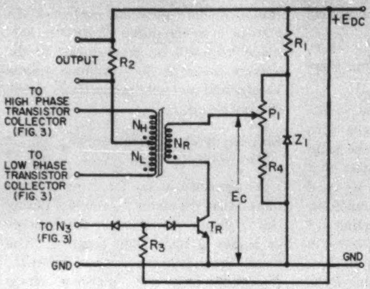


Fig. 6 (right). Undervoltage circuit



+EDC
OUTPUT
R2
R1
TO HIGH PHASE TRANSISTOR COLLECTOR (FIG. 3)
TO LOW PHASE TRANSISTOR COLLECTOR (FIG. 3)
NH
NR
NL
P1
Z1
R4
EC
TO N3 (FIG. 3)
TR
R3
GND
GND

Fig. 7 (left). Unbalanced phase sensing circuit

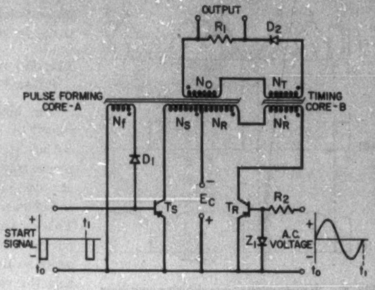


Fig. 8 (right). Fixed-time delay circuit with instantaneous reset

winding N_s constitutes part of the high-phase load that was described for the high- and low-phase sensing circuit. Therefore, the pulse that is applied to this winding during the "readout" period for the sensing circuit has a volt-second content that is proportional to the volt-second content of the system high-phase voltage. The flux change produced on the overvoltage core during this period is proportional to the average value of the highest phase voltage and inversely proportional to the system frequency.

During the next "read" period for the sensing circuit, the overvoltage core is reset by a reference pulse. To do this, the reset transistor T_R is biased to saturation by the secondary voltage of phase B. This voltage controls this transistor so that it is biased to saturation for essentially the total time of one half the period of this voltage. As a result of these considerations, it is seen that the flux change produced on the overvoltage core when it is reset is proportional to the constant voltage E_c and is inversely proportional to the frequency of the system voltages. Because both the input pulse to this circuit and the reset pulse are inversely proportional to the frequency, the comparing function that is performed on the overvoltage core is independent of frequency.

When the average value of the highest phase voltage is such that the net flux change is zero after an input pulse has been applied to the overvotage core and the core has been reset, the minimum trip condition is present. When the average value of this voltage is less than the minimum trip value, the overvoltage core saturates during its reset period. When this occurs, the current through N_R is limited by the base feed of transistor T_R. When the average value of the highest phase voltage exceeds the minimum trip voltage, the overvoltage core saturates after a time delay during a "readout" period of the sensing circuit and the high current pulses that occur as a result produce high narrow voltage pulses at the output terminals across R_2.

If the highest phase voltage is instantaneously increased from a normal value to one that is greater than the minimum trip voltage, the overvoltage core will saturate after a time delay which is inversely proportional to the difference between the average value of the alternating voltage and the minimum trip voltage (equation 19 in Appendix II). For each complete cycle of the alternating voltage there is a small net flux change in the overvoltage core. This flux change is inversely proportional to the frequency and since the number of pulses applied to the core is proportional to the frequency, the time required for the operating point of the core to reach saturation is approximately independent of frequency.

The reference voltage E_c, which is required to form the reset pulse, is provided by the Zener diode Z_1 and resistor R_1. The potentiometer P_1 and the resistor R_4 are provided for adjustment of the minimum trip voltage.

Undervoltage Protection Circuit

The function of an undervoltage circuit is to produce an alarm signal when the average value of the lowest phase voltage is less than a specified value. The circuit of Fig. 6 performs this function essentially independently of the system frequency. The circuit operates in the following manner.

During the "readout" period of the high- and low-phase sensing circuit, the low-phase pulse is applied to winding N_s of the undervoltage core. As described previously, this pulse is constant in magnitude and its width is proportional to the average value of the lowest phase voltage divided by the frequency of the system. This voltage pulse produces a proportional positive flux change on the undervoltage core. Transistor T_R is open during this period so induced current does not flow in winding N_R.

Reset of the undervoltage core occurs during the "read" period of the high- and low-phase sensing circuit. Transistor T_R is then biased to saturation for a period of time inversely proportional to the system frequency. As for the overvoltage circuit, this frequency dependence is introduced to compensate for the frequency dependence of the low-phase pulse, so, in effect, the comparing function performed on the undervoltage core is essentially independent of the system frequency. The voltage pulse applied to winding N_R during this period has a constant magnitude E_c resulting from the regulating action of Zener diode Z_1. This pulse produces a proportional negative flux change on the undervoltage core.

The turns-ratio of the undervoltage core and the value of E_s is chosen so there is no net flux change per cycle when the average value of the lowest phase voltage is equal to the specified minimum voltage. When it is higher than this value, the core will saturate in the positive direction. When this occurs the current through winding N_s is limited by the base signal on the low-phase transistor T_L of Fig. 3. When the average value of the lowest phase voltage is less than the minimum specified value, the core will saturate after a number of cycles in the negative direction during the reset period for the undervoltage circuit. An alarm signal then occurs in the form of narrow voltage pulses across resistor R_2.

Unbalanced Phase Protection Circuit

The function of an unbalanced phase protection circuit is to provide an alarm signal when the difference between the average values of the highest phase voltage of the system and the lowest is greater than a specified value. This function is performed essentially independently of frequency by the circuit of Fig. 7. The circuit operates in the following manner.

During the "readout" period of the sensing circuit the high-phase pulse is applied to winding N_H and the low phase pulse to N_L. These pulses are rectangular, of equal and constant magnitude, and have widths that are proportional to the average values of the highest and lowest phase voltages respectively divided by the frequency of the system. If the average values of the highest and lowest phase voltages are equal, the application of these two pulses will not produce a flux change on the unbalanced phase core because winding N_H and N_L are wound in opposite directions. However, if the highest phase voltage is greater than the lowest, there will be a positive change in flux that is proportional to the difference. During this period transistor T_R blocks the flow of induced current in the reset winding N_R.

During the "read" period of the high- and low-phase sensing circuit, transistor T_R is biased to saturation for a period of time inversely proportional to the system frequency. Consequently, a voltage pulse whose magnitude is E_c (a constant value determined by the combination of zener diode Z_1 and resistor R_1) is applied to winding N_R. This pulse causes a negative flux change on the unbalanced phase core of an amount that is inversely proportional to the system frequency. This frequency dependence compensates for the frequency dependence of the high- and low-phase pulses so the flux change per cycle is essentially independent of the system frequency.

It is therefore apparent that the core will remain at negative saturation so long as the difference between the average values of the highest and lowest phase voltages is proportionally less than the voltage E_c. When this difference is greater than this value, the core will become saturated in the positive direction after a number of cycles, and an alarm signal will appear in the form of narrow voltage pulses across the resistor R_3.

Fixed-Time Delay Circuit

This circuit is shown in Fig. 8. It is intended to provide an output signal a specified period of time after the application of an input signal. It will be shown in the following that the circuit is essentially independent of frequency variations over a normal aircraft frequency range and that it has a practically "instantaneous reset" characteristic.

Circuit operation is as follows. Assume initially that core A and core B have been reset to positive saturation. When an input signal is present, narrow negative voltage pulses are applied to transistor T_s. Because of the feedback produced by winding N_f, each of these pulses causes the voltage E_c to be applied to winding N_s until core A saturates in the negative direction. It should be noted that this action is independent of the shape of the input pulse. During the time that the operating point of core A is moving from positive saturation to negative saturation, a voltage

$$\frac{N_0}{N_s} E_c$$

is applied to winding N_T of core B. Consequently, there will be a small flux change on core B proportional to the maximum flux density of core A.

After core A has saturated in the negative direction, the voltage induced in winding N_f disappears and transistor T_s is biased off. Transistor T_R is then biased on for a period of time equal to one half the period of a system alternating voltage. This is accomplished by clipping a half sine wave of the alternating voltage with a Zener diode and applying this voltage to the base of T_R. Zener diode Z_1 and resistor R_2 provide this voltage pulse. There is a resulting flux change on core A during this reset period that is proportional to the constant voltage E_c and inversely proportional to the frequency. At this time a design requirement will be noted. Reactors A and B will be designed so that the magnetizing current required by N_R' will be much greater than that required by N_R. Because of this requirement all of the reset volt-seconds is absorbed by winding N_R during the reset period unless core A saturates. The core does not saturate unless the input signal is removed, and so for the present analysis, we will assume that all of the reset volt-seconds are absorbed by winding N_R. The induced current in winding N_0 is blocked during the reset period by diode $D2$ and, therfore, the flux change on core B is zero during the reset period.

The next input pulse causes core A to saturate again through winding N_s in the negative direction. During this time a voltage is applied to winding N_T of core B as described before, and, consequently, there is another small flux change in that core. This flux change is proportional to the flux change that occurred during reset of core A and is, therefore, inversely proportional to the system frequency. After core A saturates in the negative direction, the transistor T_s is biased off and the core is reset as before.

The incremental flux change produced on core B during each sensing period following the first is proportional to the constant voltage E_c and inversely proportional to the frequency. Core B will absorb a specific number of these flux changes whereupon it will saturate in the negative direction. The time delay that has occurred during the time between the first input pulse and the saturation of core B is approximately equal to the number of these pulses that is required for saturation of core B divided by the frequency. Let us say that N of these incremental flux changes are required for saturation of core B. Knowing the magnitude of these flux changes and the saturation flux of core B, we can compute N. It will then be found that N is proportional to the frequency of the alternating voltage. Therefore, the time delay is essentially independent of frequency. Actually, a frequency dependent term $1/f$ is present in the expression for the time delay. For normal frequencies this term is negligible.

TEST RESULTS ON HIGH- AND LOW-PHASE SENSING CIRCUIT AND OVERVOLTAGE CIRCUIT

A slightly modified version of the high- and low-phase sensing circuit of Fig. 3 was subjected to tests which produced

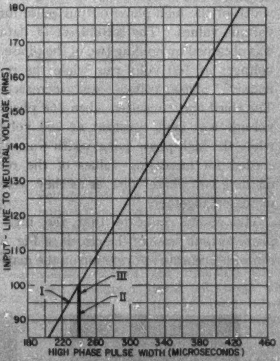


Fig. 9. Width of high-phase pulse obtained from high- and low-phase sensing circuit

I—Three-phase voltages increased together

II—Voltage of phase C is constant at 100 volts (rms). Voltages of phases A and B are increased together

III—Voltages of phases B and C are constant at 100 volts, rms. Phase A is increased

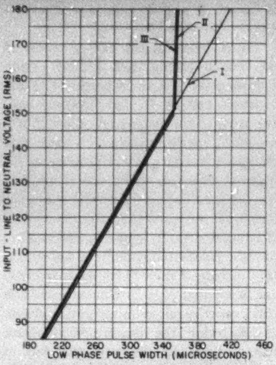


Fig. 10. Width of low-phase pulse obtained from high- and low-phase sensing circuit

I—Three-phase voltages decreased together
II—Voltage of phase C is constant at 150 volts (rms). Voltages of phases A and B are decreased together
III—Voltages of phases B and C are constant at 150 volts (rms). Phase A voltage is decreased

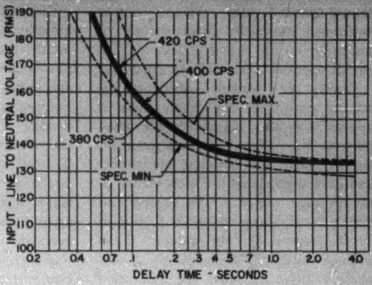


Fig. 11. Characteristics of overvoltage circuit for different system frequencies

data indicative of the accuracy of the high- and low-phase voltage sensing in a 3-phase system. High-phase sensing accuracy is indicated in Fig. 9 in which three sets of data are plotted. One set of data was obtained by increasing the three phase voltages together and recording the width of the high-phase output pulse for various values. The second set of data was obtained by holding two of the phase. voltages at 100 volts rms while the remaining phase voltage was increased. The third curve resulted when one phase voltage was kept constant at 100 volts while the high-phase pulse width was recorded for increasing values of the remaining two phase voltages as they were increased together.

The results indicate that the high-phase output pulse of the circuit of Fig. 3 contains accurate information on the magnitude of the highest phase voltage whether there are one, two, or three of these voltages which are highest. The sharp change in the slopes of the two curves that resulted when one and then two of the phase voltages were kept constant at 100 volts indicates the high sensitivity of the circuit.

The results of tests on the low-phase sensing capabilities of the circuit of Fig. 3 are shown in Fig. 10. These tests were similar to the high phase tests that have been described. It appears that the low-phase sensing is not quite as accurate as the high-phase sensing of the circuit because there is a noticeable spread between the three curves. However, it is very probable that this low-phase sensing can be improved if it is necessary for a particular specification.

The results of tests performed on the overvoltage circuit of Fig. 5 used in conjunction with the high-phase sensing circuit of Fig. 3 are shown in Fig. 11. The dashed curves represent the limits imposed by the particular specification for which the circuit was designed. The characteristics of the circuit for various frequencies of the 3-phase system are shown. The overvoltage circuit output pulse for various values of the input voltages is shown in Fig. 12. A comparison between waveforms 12(B) and 12(C) shows that the resolution of this circuit is quite high since the difference in these two waveforms is caused by a change in the high-phase voltage of less than one volt.

Combined Circuitry

When the circuits that have been described are combined to conform to the block diagram of Fig. 1, it will be found that some of the functions that were performed by a number of transistors in the separate circuits can be performed by a smaller number of transistors in the combined circuit. For example, the reset function for each of the protective circuits, which in the separate circuits requires a transistor for each circuit, requires in the combined circuit a total of only one or two transistors depending on the accuracy required. Also, the reference voltages for the circuits can be provided from a single regulated voltage supply.

Conclusion

It is felt that the following circuit concepts introduced in this paper are important:

1. The use of the memory characteristics of square-loop magnetic cores to store information as to the magnitudes of the average values of two or more alternating voltages that are out of time phase so that this information may be obtained or "read out" at any convenient time. Since pulses that have volt-second contents proportional to the average value of each of the sensed voltages may be obtained simultaneously, their volt-second content may be compared and the high-phase and low-phase pulses measured.
2. The use of the inherent capability

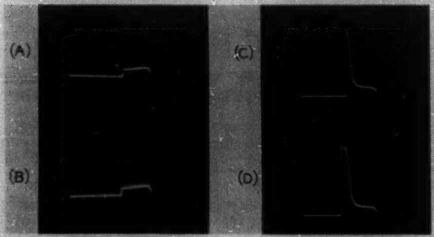


Fig. 12. Output pulse of overvoltage circuit for different values of system voltages

A—All phases = 120 volts
B—Phases A and C = 120 volts. Phase B = 134 volts
C—Phases A and C = 120 volts. Phase B = 135 volts
D—Phases A and C = 120 volts. Phase B = 155 volts

of square-loop magnetic cores to sense the total amount of volt seconds underneath the waveforms of the sensed voltages or the true average value of each of these voltages. Consequently, the average values of the sensed voltages are sensed regardless of their harmonic content.

3. The use of reactors using square-loop magnetic cores to provide a time delay that is inversely proportional to the difference between the average value of the sensed voltage and a "minimum trip" voltage which is determined by the average value of the reference pulse. The width of the reference pulse is made proportional to the period of the sensed voltage while its magnitude is regulated by a Zener diode so that the time delay produced by the core is essentially independent of the frequency of the sensed voltage.

4. The use of the memory characteristics of reactors using square-loop magnetic core material to compare the volt-second content of two voltage pulses. This function is also performed independent of frequency by making the width of the reference pulse proportional to the period of the sensed voltage.

The use of transistors or other devices as switches is necessary to perform some of the foregoing functions. In other cases, the use of transistors offers certain advantages such as; less power dissipation, fewer adjustments, smaller and lighter components, straightforward design, greater accuracy, less sensitivity to variations in parameter characteristics, etc. The circuitry described in this report has been designed with the intention of providing all of the possible advantages of this type of circuitry so as to demonstrate its full capability.

The design of transistor-magnetic circuitry to meet a particular specification may not require all of the advantages that have been mentioned and consequently the circuitry may be simplified. The decision as to how the circuit is to be designed to meet a particular specification can only be made after a thorough analysis of all of the factors involved.

Appendix I

Volt-Second Content of Output Pulse of Pulse Forming Circuit

For the first part of this analysis, refer to Fig. 2. During the sensing period when switch T_s is closed and switch T_R is open, the flux change of the pulse forming core is

$$\Delta\phi_s=\frac{1}{N_s}\int_0^{\frac{\pi}{w}} k_t e\,dt \tag{1}$$

where w is the angular frequency of the input voltage e and k_t is the ratio of transformation N_p/N_s of the input transformer. Since the average value of the input voltage is defined as

$$E_{avg}=\frac{w}{\pi}\int_0^{\frac{\pi}{w}} e\,dt \tag{2}$$

Equation 1 can be written as

$$\Delta\phi_s=\frac{k_t}{N_s}\frac{\pi}{w}E_{avg} \tag{3}$$

During the reset period when switch T_s is open and switch T_R is closed, the flux change on the core is

$$\Delta\phi_R=\frac{1}{N_R}\int^{\Delta t} E_{dc}\quad dt=\frac{1}{N_R}E_{dc}\Delta t \tag{4}$$

where Δt is the length of time between the closing of switch T_R and the saturation of the core.

Since this flux change is in the opposite direction to the flux change during the sensing period, we can solve for Δt by setting equation 4 equal to equation 3

$$\Delta t=\frac{N_R}{N_s}\frac{\pi}{w}\frac{k_t E_{avg}}{E_{dc}} \tag{5}$$

During the reset period, a voltage is induced across winding N_0 that is equal to:

$$E_o=\frac{N_c}{N_R}E_{dc} \tag{6}$$

The duration of this output voltage pulse is given by equation 5. If this voltage pulse is applied to the base of a transistor so that the transistor is saturated throughout the duration of the pulse and the load for this circuit is placed in series with the collector and the voltage E_{dc} the volt-second content of the voltage pulse applied to this load is

$$(VT)_o=\frac{N_R k_t}{2N_s}\frac{E_{avg}}{f} \tag{7}$$

It should be noted that it is not necessary to have a sine-wave input voltage for this equation to be applicable.

As a result of the foregoing analysis, it is apparent that for the 3-phase circuit of Fig. 3, the volt-second content of the voltage pulses applied to the high-phase load and the low-phase load are respectively

$$(VT)_h=\frac{N_R k_t}{2N_s}\frac{E_H}{f} \tag{8}$$

$$(VT)_L=\frac{N_R k_t}{2N_s}\frac{E_L}{f} \tag{9}$$

where E_H and E_L are respectively the average value of the high-phase input voltage and the average value of the low-phase input voltage.

Appendix II

Analysis of Overvoltage Circuit with Inverse Time Delay

The circuit that will be analyzed is that of Fig. 5. Symbols that refer to the reactor in this circuit will be primed. The high-phase output pulse of the high- and low-phase sensing circuit constitutes the input signal; the output signal is obtained in the form of high, narrow voltage pulses across resistance R_3. The reset voltage or the reference voltage E_c is a constant voltage regulated by the combination of the Zener diode Z_1 and the resistance R_1. The sensing and reset periods that will be referred to are those of the overvoltage circuit, not of the sensing circuit. In other words, the sensing period of the overvoltage circuit occurs during the same time as the reset period of the sensing circuit. Transistor T_R is biased on during the positive half cycles of the alternating voltage for a period of time approximately equal to one half the period of this voltage.

Assume initially that the overvoltage core has been reset by E_c to saturation. During the next sensing period, the high-phase output pulse of the sensing circuit which contains an amount of volt-seconds given by equation 8 of appendix I is applied to winding N_s'.

The flux change which is produced by this pulse is

$$\Delta\phi_s'=\frac{N_R k_t}{2N_s'N_s}\frac{E_H}{f} \tag{10}$$

where E_H is the average value of the highest phase voltage.

During the next reset period the voltage E_c is applied to winding N_R' for one half the period of the alternating voltage. The flux change produced during this reset period is, therefore

$$\Delta\phi_R'=\frac{1}{2N_R'}\frac{E_c}{f} \tag{11}$$

The minimum trip value of the alternating voltage will be defined as the voltage that exists when the change in flux on the overvoltage core during its sensing period is equal to the change in flux during its reset period.

We then have

$$\Delta\phi_s'=\Delta\phi_R' \tag{12}$$

Substituting from equations 1 and 2 we have

$$E_H^{o}=\frac{N_s'N_s}{N_R'N_R}\frac{E_c}{k_t} \tag{13}$$

where E_H^{o} is the value of the alternating voltage that has been defined as the minimum trip voltage and will theoretically produce an output signal from the overvoltage circuit after an infinite time delay. Therefore, we see that when the alternating voltage exceeds the minimum trip voltage, an output signal will occur after a time delay. In the following it will be shown that this time delay is essentially inversely proportional to the difference between the high-phase value of the alternating voltage and the minimum trip value of this voltage.

The net change in flux produced by one sensing period and one reset period (in other words, produced by one complete cycle of the alternating voltage) is from equations 1 and 2

$$\Delta\phi=\frac{N_R}{2N_s'N_s}\frac{k_t E_H}{f}-\frac{1}{2N_R'}\frac{E_c}{f} \tag{14}$$

After some manipulation, this becomes

$$\Delta\phi=\frac{1}{f}[E_H-E_H^{\circ}]\frac{N_RK_t}{2N_s'N_s} \tag{15}$$

The overvoltage core can absorb a specific number of these net flux changes. This number N can be determined from the following equation

$$\frac{N}{f}[E_H-E_H^{\circ}]\frac{N_Rk_t}{2N_s'N_s}=2\phi_s'-\frac{N_Rk_t}{2N_s'N_s}\frac{E_c}{f} \tag{16}$$

The amount of time that has elapsed between the time when the input voltage was instantaneously increased to a value that is greater than the minimum trip voltage and the time when the overvoltage core saturates during the sensing period is

$$T=\frac{N}{f}+\frac{1}{2f} \tag{17}$$

Solving for N from equation 7 we obtain

$$N=\frac{\dfrac{4N_s'N_s}{N_Rk_t}f\phi_s'-E_c}{[E_H-E_H^{\circ}]} \tag{18}$$

Substituting this expression in equation 17 we obtain as the expression for the time delay

$$T=\frac{\dfrac{4N_s'N_s}{N_Rk_t}\phi'_s-\dfrac{E_c}{f}}{[E_H-E_H^{\circ}]}+\frac{1}{2f} \tag{19}$$

Manufacture of Composite Cable Sheath for Telephone Exchange Cable

D. A. HUGHES
NONMEMBER AIEE

IN VARYING degrees, the insulated conductors used in telephone cables require protection from mechanical damage, moisture, and electrical disturbance. From very early in the history of telephone circuits, lead sheath provided this protection.

During and immediately following the Second World War, lead was in critically short supply. In order that telephone expansion not be limited by the lead shortage, the Bell Telephone Laboratories and the Western Electric Company developed cable sheaths which contained no lead.[1–4] A cutaway view of each of the two general types of replacing sheath is shown in Fig. 1.

To obtain sufficient sheath conductivity, aluminum proved to be the most economically usable material. Where maximum moisture protection is required, as for paper insulation, a moisture barrier of soldered steel is placed over the aluminum. For over-all protection against mechanical damage, corrosion, and moisture, an asphalt base cement is applied and a tube of high-molecular-weight polyethylene is extruded over the assembly of core, aluminum, and steel.

Polyethylene-insulated conductor, which requires less moisture protection than paper ribbon or pulp, is sheathed with aluminum and polyethylene (Alpeth sheath) in sizes from 0.22–2.38-inch core diameter. For core diameters of 0.68 inch and larger, the aluminum strip is corrugated to maintain adequate flexibility.

Paper- or pulp-insulated conductor is sheathed with aluminum, steel, and polyethylene (Stalpeth sheath) in sizes from 0.40–2.38-inch core diameter. The steel strip is always corrugated; the aluminum strip is corrugated on 0.90-inch-diameter cores and larger. This corrugation permits easier flexing of the cable with less damage to the metal strip. The range of sizes of composite sheaths is shown in Table I. In the region between 0.68-inch and 0.89-inch core diameter, the corrugated steel of Stalpeth sheath supports the aluminum sufficiently to make it unnecessary to corrugate the aluminum.

The aluminum strip used in these sheaths is 8/1000 inch thick. The steel is 7/1000 inch thick and tin plated. The polyethylene thickness is 0.045–0.110 inch. The solder is 30% tin, 70% lead. The rubber tape used over polyethylene-insulated conductor as core-to-sheath insulation is 15/1000 inch thick with 1/1000 inch of Mylar on the outer face.

Manufacture of Aluminum-Polyethylene Sheath (Alpeth)

Polyethylene-insulated conductor is twisted into pairs and subsequently into units and cable cores. Throughout the process, the conductor is exposed to normal room temperature and humidity. The completed core is fed into the Alpeth sheathing line.

The Alpeth sheathing operation proceeds as follows: First, the Mylar-faced rubber tape is applied longitudinally to give greater core-to-sheath breakdown strength and to protect the conductor

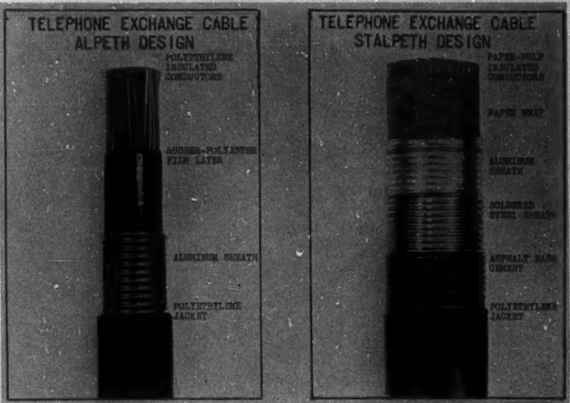


Fig. 1. Cutaway view of each general type of composite cable sheath

---

Paper 59-792, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 11, 1959; made available for printing May 4, 1959.

D. A. Hughes is with the Western Electric Company, Kearny, N. J.

Table I. Range of Sizes of Composite Sheaths Manufactured by Western Electric Company

Conductor Insulation	Type of Sheath		Range of Sizes Made, Core Diameters in Inches			
	General	Subdivision	0.22–0.39	0.40–0.67	0.68–0.89	0.90–2.38
Polyethylene	aluminum and polyethylene (Alpeth)	uncorrugated aluminum	X	X		
		corrugated aluminum			X	X
Paper pulp	aluminum, steel, and polyethylene (Stalpeth)	corrugated steel and uncorrugated aluminum		X	X	
		corrugated steel and corrugated aluminum				X

insulation from the subsequent heat of extrusion. Second, the aluminum strip, flat or corrugated, is formed around the core. Third, the polyethylene is extruded over the aluminum, cooled, and dried.

The thickness of the polyethylene jacket depends on the ratio of the volume of polyethylene extruded to the length and diameter of core over which it is extruded. Hence, for a given core, extruder screw, die setup, and temperature distribution, the output volume for each speed is measurable. The tractor capstan speed should be adjusted to maintain a constant ratio between output volume and line speed. In this process, the adjustment is made automatically.

The screw speed signal is used as the basis of comparison. When the screw speed is increased, its speed proportional signal is compared with a modified signal from the tractor capstan drive. The interaction of these signals increases the drive speed to maintain approximately constant, the ratio between extruder output volume and line speed. Hence, as the screw speed is changed, approximately constant thickness is maintained automatically.

The concentricity of the polyethylene jacket depends upon the relative position of core tube and die, and upon the temperature and pressure distribution between tube and die.

Inasmuch as the extrusion process may not be stopped while a length of cable is in the line, it is necessary to provide accumulators for the rubber tape and the aluminum tape for the time required to splice each. When the coil of rubber tape approaches a predetermined distance from runout, the speed of pay-off is increased to approximately double the line speed so that an excess of tape is forced into an accumulator. When the coil of tape runs out, there is sufficient length of tape in the accumulator so that the end of the tape can be held and cemented to the start of the new coil while the accumulator continues to supply the line. The accumulation of aluminum tape to permit the welding of the end of the expiring coil to the beginning of the new coil while the accumulator supplies the line is similar. Dual cable core supply and take-up positions are provided also to avoid stopping the extruder.

A capacitance bridge whose recording chart is graduated in thousandths of an inch of polyethylene jacket thickness, records the jacket thickness at octants around the cable.[5–7]

In addition to its use as an inspection record, this chart is used by the operator to check concentricity, roundness of core, and average thickness. He will then make appropriate adjustments if necessary and recheck.

The finished product is reeled directly upon the shipping reels at the end of the line. Before shipment the cable is tested for core-to-sheath breakdown, conductor resistance, short, cross and open circuits, dielectric strength, mutual capacitance, and capacitance unbalance. Some of these items are checked on a statistical sampling basis.

Manufacture of Aluminum-Steel-Polyethylene Sheath (Stalpeth)

Paper-pulp-insulated conductor is twisted into pairs and subsequently into units and cores at normal room temperature and humidity.[8] Paper of 0.005-inch thickness for core-to-sheath insulating strength is applied in double thickness in the cabling operation.

The completed core is dried by the passage of direct current through its conductors while in a vacuum drying dome.[9] When the core is dry, it is removed from the drying dome and placed immediately in a hot room, 130 F (degree Fahrenheit), to prevent excessive moisture absorption by the pulp. The pay-off stands from which the core is supplied to the steel and aluminum sheathing line are inside the hot room. The metal covering line, with the exception of the core pay-off stands and process reel take-up stand, is shown in Fig. 2.

For core diameters smaller than 0.90 inch, the steel tape is corrugated but the aluminum tape is flat. The flat aluminum tape is formed longitudinally around the core as the core enters the tube of the belt forming machine. This machine forms the corrugated steel tape longitudinally around the core and the aluminum.

When the core diameter is larger than 0.90 inch, both aluminum and steel tapes are corrugated. The steel tape is corrugated first and then the aluminum tape is corrugated so that its corrugations match and nest in those of the steel. The steel

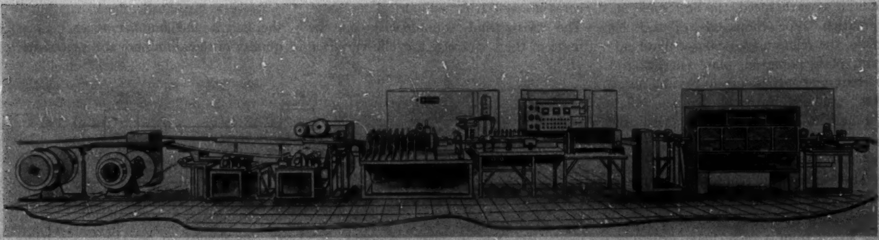


Fig. 2. Metal covering line. Artist's sketch showing cable core entering line from left, aluminum strip pay-off, steel strip pay-off, steel corrugator, aluminum corrugator, belt-forming machine, r-f generator, soldering position, operator's control panel, twist corrector, drier, tractor capstan, open and weak seam detector, and metal-sheathed core leaving line to right

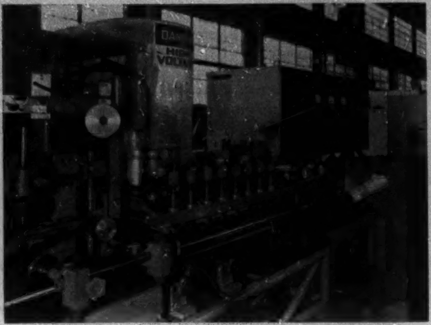


Fig. 3. Soldering position showing formed metal-covered core entering at left, flattened solder feed, r-f output current transformer, pressure-shoe assembly of soldering and cooling shoes, and operator's control panel

tape, with the aluminum tape lying on top of it and its corrugations nesting in (in mesh with) those of the steel, is fed into the belt forming machine.[10] Both tapes are formed together against a tube around the core and a strip of paper is inserted under the lap to prevent damage of the core by heat when the seam is soldered. A forming die completes the forming of the metals around the core.

Flattened rosin core solder strip is inserted in the steel lap before it proceeds beneath the r-f (radio-frequency) inductor. The soldering position, which includes the solder inserting tool and the inductor, is shown in Fig. 3.

The area where heating is desired is the strip, approximately 1/2 inch wide, which includes the steel edges being soldered and the included solder. Under the single-turn hairpin shape inductor, heat is induced in the lap to melt the solder, fill the lap, and wet the steel edges with molten solder.

It has been observed that best soldering is attained when the steel is heated quickly to above 600 F and quenched immediately.

To control the heating in the soldering operation automatically, the energy level of the generator was adjusted manually to obtain, at each speed, the best soldering condition. For thousands of reels of cable, the relation of oscillator plate current, plate potential, plate energy, grid current, and output current to cable speed was studied. It was learned that, throughout a wide range of speed, the oscillator energy input should be proportional to the cable speed.

The designer and builder of the present 50-kw r-f generators developed a control which proportions the oscillator input to the signal from a tachometer generator. Over the required range of speed and energy levels, this response is a linear proportion.

The tachometer generator used for the cable speed proportional signal is driven by a sheave geared to the corrugations of the completed steel sheath as it leaves the tractor capstan. This signal is fed through an operator-adjusted potentiometer to the input to a 2-stage magnetic amplifier. The amplified signal feeds the saturating winding of a saturable reactor. This reactor controls the a-c input to the generator rectifier's high-voltage transformer primary.

The over-all effect of this control is to produce the proper heating effect for any cable speed while allowing the operator to adjust quickly if the cable assembly or coil distance is changed. Fig. 4 shows a block diagram of this control loop together with the solder flattening control loop which is described later.

In order to obtain uniform heating of the seam being soldered at a given speed and energy level, it is necessary to have steady coupling of the inductor with the seam. To maintain constant coupling of inductor to seam when the seam varies in position, the inductor is designed to follow the seam's position and keep a constant, adjustable, distance from the seam.[11]

This feature is required because cable cores are not truly cylindrical in shape. The twists and countertwists in the units of the core cause a cyclic variation in contour of any longitudinal section of the core. The aluminum and steel, when formed around the core, follow this contour.

At one point in the development, the use of a much lower frequency for induction heating was investigated. In some respects, the rotating source of energy is simpler and more rugged than an electronic r-f generator.

A rotating source of 10-kc energy was used to heat samples of the sheathed cable. This lower frequency is inherently more difficult to couple than the 450 kc used in previous work. Because of the coupling difficulty, a more massive output current transformer and inductor were required for comparable heating efficiency.

If 10-kc energy were used, the more massive inductor would complicate the soldering position design. The pressure shoes would have to be larger and less convenient to control. The feature of following the seam at a constant distance would be more costly to provide and more difficult to control. In present design, the 450-kc transformer and inductor are made into one structure, pivoted, and guided for ease of adjustment and to maintain constant distance between inductor and work.

The instant that the seam leaves the inductor, it is quenched with water to product the soldered seam.[12] Care must be exercised to prevent relative motion of the steel edges during cooling.

Air blasts and vacuum remove the water from the steel sheath. The sheath is stressed by rollers after it leaves the tractor capstan. During this stressing, a weak seam will open and make contact with electrodes provided and stop the operation until the operator repairs the soldered seam. The aluminum and steel sheathed core is reeled on a process reel to permit core-to-sheath insulation resistance and breakdown tests before the polyethylene is applied.

Process inspectors check continually the height and number of corrugation, quality of the aluminum and plated steel,

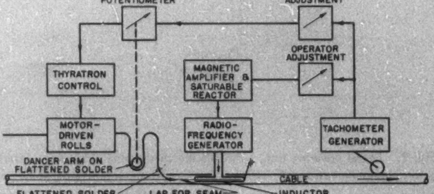


Fig. 4. Block diagram of control loops for solder flattening and heat regulation

Vol
Pt.
195

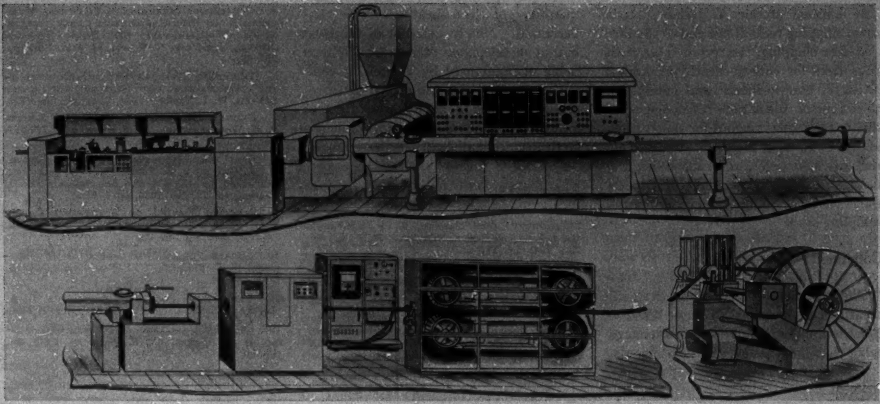


Fig. 5. Polyethylene jacketing line. Artist's sketch showing metal-sheathed core entering line from left, rounding rollers, asphalt base cement flooding tank, extruder, operator's control panel, pressure cooling tube, drier, sheath thickness scanner and bridge, tractor capstan, and take-up winding completed cable on shipping reel

register, and continuity of the soldered seam.

Several factors are checked continuously by gaging methods built into the manufacturing equipment. The steel strip is checked by a photoelectric relay which stops the process if a hole of about 1/100-square-inch area or larger is detected. The width of the steel strip is also checked and the process stopped in the event of narrow strip. A break in the continuity of, or the motion of, the solder wire will also stop the line. Failure of the r-f supply or a defective soldered seam will also stop the process.

When one of these events causes the process to stop, one of the signal lights at the top of the operator's control panel indicates which event caused the stop.

To permit the use of standard round rosin core solder, a solder flattening mill has been provided on each line. The round wire is flattened to the volume required by the action of rolls whose temperature is thermostatically controlled and whose speed is proportioned to that of the cable in the line.

For this control the same speed proportional signal as used for the r-f heating control, is fed through an operator-adjusted rheostat and through a dancer arm adjusted potentiometer to a Thyratron speed control.

The roll of the dancer arm rides on a loop of the solder as it leaves the flattening mill. An additional control feature is the introduction of a booster signal when the dancer roll nears the top of its travel. If the dancer roll were permitted to reach the top, the flattened wire would be taut and likely to break. The booster signal accelerates the mill an adjustable added speed so that the dancer arm drops to the middle of its travel and removes the booster signal. At slow speeds, this cycling will continue until the cable speed signal is high enough to cause the mill to move in concert with the line.

The Thyratron control proportions the speed of the motor drive to flatten and feed precisely enough solder for the cable being sheathed.

The metal-sheathed core is paid into the polyethylene jacketing line which is shown in Fig. 5. Since the sheath has been somewhat flattened during stressing, first in the vertical axis and then in the horizontal axis, and to some degree deformed in reeling and unreeling, a sequence of forming rolls is provided. These rolls restore the original roundness of the metal sheath. Asphalt base cement is applied to fill the valleys of the corrugations and provide an adhesive between the steel and the polyethylene. The polyethylene jacket is extruded over the cement and cooled under pressure in a long tube containing water at room temperature or below. Fig. 6 is a cutaway view of the extruder head showing the feature of using vacuum to eliminate voids between the metal sheath and the polyethylene jacket. A series of high-velocity air jets dries the polyethylene. The thickness at eight successive octants is recorded by a capacitance bridge.

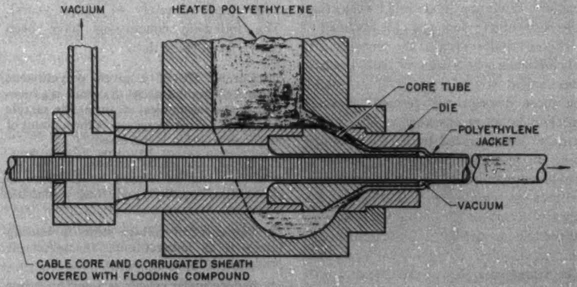

VACUUM
HEATED POLYETHYLENE
CORE TUBE
DIE
POLYETHYLENE JACKET
VACUUM
CABLE CORE AND CORRUGATED SHEATH COVERED WITH FLOODING COMPOUND

Fig. 6. Polyethylene extruder cutaway sketch showing application of polyethylene jacket to metal-sheathed core with vacuum eliminating voids between metal and jacket

The finished product is reeled directly upon shipping reels at the end of the line.

Before shipping, the reel of cable must pass tests for core-to-sheath breakdown strength, core-to-sheath and conductor-to-conductor insulation resistance, conductor resistance, ringout for short, cross, and open circuits, dielectric strength, a-c mutual capacitance, and capacitance unbalance. As is the case with Alpeth, some of these items are checked on a sampling basis.

Conclusion

The quantity of Alpeth and Stalpeth sheath manufactured by the Western Electric Company for telephone exchange cable has increased greatly while the quantity of lead-sheathed cable has decreased. Ninety per cent of the total output of telephone exchange cable now being manufactured by Western is composite sheath. Use has shown these replacing sheaths to be the equivalent of lead in life and more economical to manufacture and install.

References

1. Alpeth Cable Sheath, **R. P. Ashbaugh.** *Bell Laboratories Record*, New York, N. Y., vol. 26, Nov. 1948, pp. 441–44.

2. Stalpeth Cable Sheath, **R. P. Ashbaugh.** *Ibid.*, vol. 29, Aug. 1951, pp. 353–55.

3. Electric Cable Sheathing, **H. G. Johnstone.** *U. S. Patent no. 2,589,700*, Mar. 18, 1952.

4. Bell System Cable Sheath Problems and Designs, **F. W. Horn, R. B. Ramsey.** *AIEE Transactions*, vol. 70, part II, 1951, pp. 1811–16.

5. Continuous Incremental Thickness Measurement of Non-Conductive Cable Sheath, **B. M. Wojciechowski.** *Bell System Technical Journal*, New York, N. Y., vol. 33, Mar. 1954, pp. 353–68.

6. Thickness Measurement and Control in the Manufacture of Polyethylene Cable Sheath, **W. T. Eppler.** *Ibid.*, May, pp. 554–77.

7. Grounded Capacitance Measuring System, **B. M. Wojciechowski.** *U. S. Patent no. 2,721,975*, Oct. 25, 1955.

8. Pulp Insulation for Telephone Cables, **H. G. Walker, L. S. Ford.** *AIEE Transactions*, Mar. 1933, pp. 133–41; also *Bell System Technical Journal*, vol. 12, Jan. 1933, pp. 1–21.

9. Electric Drying of Telephone Cable, **L. G. Wade.** *Bell System Technical Journal*, vol. 19, Apr. 1940, pp. 209–20.

10. Sheath Forming Apparatus, **E. W. Reynolds, H. C. Slechta.** *U. S. Patent no. 2,764,214*, Sept. 25, 1956.

11. Cable Seam Soldering Apparatus, **D. A. Hughes.** *U. S. Patent no. 2,758,189*, Aug. 7, 1956.

12. Soldering Apparatus for Cable Sheathing, **D. A. Hughes.** *U. S. Patent no. 2,801,316*, July 30, 1957.

A Study of the Effects of Corona on Polyethylene

EUGENE J. McMAHON
ASSOCIATE MEMBER AIEE

DANIEL E. MALONEY
NONMEMBER AIEE

JOSEPH R. PERKINS
ASSOCIATE MEMBER AIEE

THE DELETERIOUS effect of corona on the insulation of wires and cables subjected to high voltages has long been recognized. This basic work was undertaken to investigate the effects of corona on commercial Fawcett-type polyethylenes (branched high-pressure free-radical polymers) used as insulation in high-voltage (HV) applications. Two questions were to be answered. What are the factors involved in the design of an HV cable? What are the criteria for selecting the best polyethylene resin for this use?

These studies revealed the existence of a phenomenon which we now call "electromechanical stress cracking." It might be more descriptively termed "corona-mechanical stress cracking." This effect can be demonstrated by carefully wrapping an insulated wire spirally about a grounded mandrel and applying sufficient voltage to cause corona. The failure always occurs in the insulation which is under tension and not at a point adjacent to the mandrel as would be expected. The many variables in the mandrel test are difficult to control. A simpler test method was devised in which a film sample is stretched to give essentially uniaxial strain. Then an electrical stress is applied through the film at right angles to the mechanical strain. Relative corona life by this method correlates with that from the mandrel test. Also the stretched film arrangement allowed the use of direct electrode voltage endurance methods to study the effects of corona on insulating materials.

The investigation has involved to date approximately 400,000 sample exposure hours (approximately 44 sample-years). The following conclusions have been drawn from these data.

1. Corona life of a given polyethylene under a constant amount of corona is a function of the mechanical strain in the sample exposed. Life is decreased as the amount of mechanical strain is increased.

2. Initial mechanical stresses within the insulation can be relaxed by exposure to elevated temperatures and, thus, corona life extended.

3. The effect of voltage stress in causing failure is an indirect one; the corona it produces is the active agent.

4. Failure under corona is a surface effect and not an internal one. In the case of voids in the body of an insulation corona attack is on the inside surface of the void.

5. Corona must be suppressed to get indefinitely long life under voltage. Furthermore, corona-free electrical stresses on polyethylene insulation can be surprisingly high and still result in indefinitely long life.

6. When samples are elongated and exposed to corona, failure follows the development of surface cracks perpendicular to the direction of elongation. In an unelongated sample, the mechanism of failure is radically different; failure occurs because of progressive erosion of the insulation surface by ion bombardment which leads to decreased insulation thickness.

7. For a given voltage level causing corona, corona life of specimens in air is a strong function of the humidity. A relative humidity (RH) of 95 to 100% gives approximately 15 times longer life than dry air. The apparent reason for this somewhat surprising result is the development of a semiconducting surface layer under humid conditions which grades the fields and diminishes the corona intensity.

8. At a given voltage level, nitrogen at atmospheric pressure results in longer corona life than air. Again increased humidity augments life considerably. Corona life under our test conditions of 200 volts per mil and 50% elongation was considerably longer under 100% RH nitrogen than any other atmosphere. Again the establishment of a semiconducting layer during initial stages of corona bombardment essentially extinguished subsequent corona.

9. At a given voltage level, an environment of carbon dioxide at atmospheric pressure gave results which were again a function of humidity and were intermediate between air and nitrogen.

Paper 59-800, recommended by the AIEE Electrical Insulation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 20, 1959; made available for printing May 15, 1959.

Eugene J. McMahon, Daniel E. Maloney, and Joseph R. Perkins are with E. I. du Pont de Nemours & Company, Inc., Wilmington, Del.

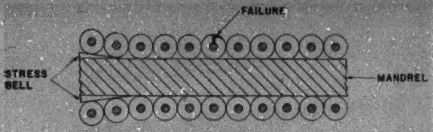


Fig. 1. Cross section of a typical failure

10. The addition of small amounts of carbon black to the polyethylene to improve resistance to weathering decreases corona life.

Implications for the Manufacturer and Users of Cables

The work described here is primarily basic research and the test specimens bear little physical resemblance to cable insulation. However, the data obtained and the conclusions drawn are directly applicable to the design and manufacture of cable. What does this information mean in cable manufacture and use?

1. Cables should be designed to be free from corona under service conditions. This means absence of either internal voids or space where corona can occur next to strand or shield.

2. The existence of the electromechanical stress-crack phenomenon means that: (*a*) Fabrication by the cable manufacturer should be carried out to give an absolute minimum of residual strain, so that the insulation will have maximum resistance to any possible corona. (*b*) The user might take steps to relax the stresses caused by bends and installation handling. This can be done by heating the cable by conductor resistance losses prior to voltage application. Even if this is not done, however, serious damage is not likely to occur since both time and temperature lead to relaxation.

3. Black jacket stock should not generally be considered as a primary insulation.

4. The improved life in nitrogen as compared with air suggests advantages for nitrogen pressure-type cables for the higher voltage levels. The use of moisture could also be considered if investigation shows that its presence is not harmful to the metal elements of the cable.

Background for Study

In view of the growing demand for plastics as insulating materials in HV applications, it became advisable to make a study of the corona life of these materials. Insulation is often subjected to a mechanical stress in addition to the electrical stress. This is particularly true in HV cable applications where the cable may be bent sharply, or mechanical stresses may be introduced in the insulation when 3-phase cables are twisted during construction. To the best of our knowledge, systematic measurements have never before been made on the effect of combined mechanical and electrical stresses. "Corona cutting" has been observed on rubber insulation when it is subjected to corona discharge. The cracking is accelerated when the rubber is elongated. The damage, in the case of rubber, is a function of ozone degradation, a reaction involving oxidation of the rubber molecule. This discussion will concern Alathon* polyethylene resin, a nonpolar plastic which does not fail by the same mechanism as rubber.

A series of experiments will be described in which the effects of applying both mechanical and electrical stresses, or electromechanical stress cracking, were studied. In this presentation, the concern is with voltages slightly above the corona starting voltage for the electrode system employed. There is a significant reduction in the life of the insulation due to corona.

Corona has been loosely defined as a luminous discharge. In this discussion, corona is defined as any electrically-detectable field-intensified ionization that does not result immediately in complete breakdown of the insulation-electrode system in which it occurs.

Test Data

Original Experiments

Early measurements to determine if there really is an effect of combined mechanical and electrical stresses were made on covered wire specimens wrapped on mandrels. The mandrels were 3/16-inch diameter stainless-steel rods approximately 6 inches long. Small electric stress bells were wound of plastic insulating tape leaving a $1^1/_2$-inch long section of uncovered steel on which to wrap the wire. Six mandrels were prepared and wrapped with 12 turns of no. *14* wire covered with polyethylene insulation having a wall thickness of 0.034 inch. Three of the wrapped mandrel samples were exposed to an average electrical stress of 200 volts per mil by applying 6,800 volts alternating current between the mandrel and the conductor. Corona was visible about the coils. The other three wrapped mandrels were not electrified but were stored nearby in the laboratory. The failures of the electrified specimens took place after 18, 26, and 27 hours. The damage was on the outside of the coils approximately at the point of greatest tension, not next to the mandrel which was the shortest path, as shown in Fig. 1. The three mandrels which were not exposed to the electrical stress, but had been under mechanical stress for one week (170 hours), showed no cracks.

For comparison, straight lengths of the same polyethylene-covered wire were loosely wrapped with tin foil to form an outer electrode. The same electric stress of 200 volts per mil was applied. The test was discontinued when no failures had occurred at 100 hours, compared with 18, 26, and 27 hours for the mechanically stressed samples.

This experiment was repeated many times with substantially the same results. Thus, preliminary proof was obtained of electromechanical stress cracking of polyethylene under the influence of corona.

In order to obtain reproducible data, it was found necessary to exercise care in winding the samples on mandrels. Good results were not obtained until the winding tool shown in Fig. 2 was built. With this tool, the wire was wrapped to a predetermined tension without kinking or twisting. No space was left between adjacent turns.

When the specimens were electrified at 200 volts per mil, a soft corona glow was seen around the coil when it was observed in a darkened room. No sharp point discharges were visible. The corona starting gradient was 80 volts per mil.

Fig. 2. Mandrel wrapping tool

---

* Registered trade-mark of E. I. du Pont de Nemours & Company, Inc.

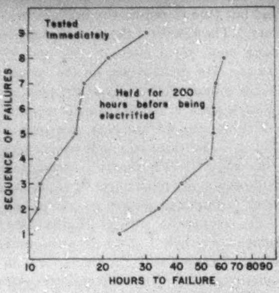


Fig. 3. Effect of mechanical stress relaxation; measurements made with electric stress of 200 volts/mil and mechanical elongation of 50% (initial)

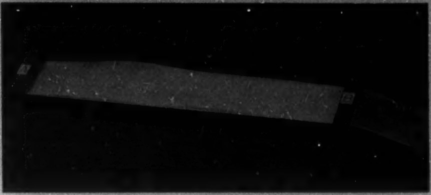


Fig. 4. Modified corona test electrode

A test rack was built to accommodate ten samples for simultaneous tests. A hood removed the ozone.

Effects of Mechanical Stress Relaxation

The effect of mechanical stress relaxation of the insulation on copper wire was investigated by wrapping 20 mandrels as previously described. Ten of these mandrels were immediately subjected to an electrical stress of 200 volts per mil. The remaining ten wrapped mandrels were held at room temperature for 200 hours before being electrified. The results in Fig. 3 show that the time to failure was longer for the specimens for which the insulating polyethylene was permitted to relax mechanically. For the latter specimens, part of the mechanical stress had been relieved.

Measurements on Films

In a mandrel test, it is very difficult to measure the amount of elongation to which a wire covering is subjected when it is bent since the forces may be triaxial. Furthermore, the amount of elongation varies throughout the cross section of the wire covering and the electrode geometry is complex; therefore, a film sample was adopted for future studies. Polyethylene films can be prepared and strained in such a fashion that the mechanical stresses are nearly uniaxial and reasonably constant throughout the cross section of the sample.

Prior to the corona studies on the mandrel wrapped samples, work was done on a direct electrode method for corona degradation measurements on insulating materials in co-operation with the ASTM (American Society for Testing Materials). The method used a rod-to-plane electrode system. The flat-plate electrode used for the ASTM study was modified. Clamps were provided so that the film could be elongated and secured under stress for the duration of the test.

Fig. 4 shows the modified electrode. The film, 2 inches wide and 14 inches long, is first secured by clamp *A* and passed through clamp *B* which is loose. The assembly is placed in a stretching rack in which the film is elongated slowly (1 inch per minute) and uniformly. The amount of stress and the elongation in the film are recorded.

The film thickness is measured and clamp *B* is tightened before the film is released from the grip. In calculating the electric stress in volts per mil, the thickness of the elongated film is used.

The lower flat-plate electrode is placed in the electrode holder (Fig. 5). The upper electrodes are constructed of stainless steel 1/2 inch in diameter and have a 1/16-inch radius on the edge. Each electrode holder is equipped with five parallel test electrodes. Each test bay holds two such electrode holders so that ten points are exposed simultaneously. Ten bays are used.

The current that flows at the time of failure is limited to 40 ma (milliamperes) by a limiting resistance network so that a large hole is not made in the sample when a test point fails. Large holes would relieve the mechanical stress on the film. An individual fusing system (Fig. 6) was devised to disconnect each individual test point automatically at the time of its failure. Current flowing from the electrode at the instant of failure passes through the voltage dropping resistance to the fuse wire. This causes a voltage difference between the chisel point and the fuse wire which gives rise to an arc burning the fuse wire which is under tension. Before the fuse wire parts, however, current passes to ground through the relay which sends a pulse to a strip chart recorder to mark the time of the failure.

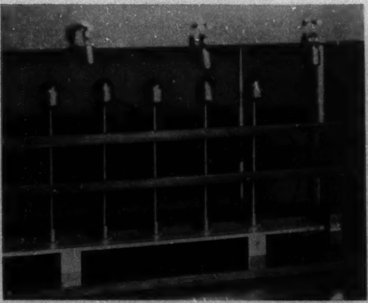


Fig. 5. Electrode assembly

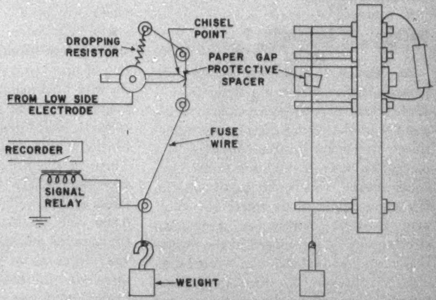


Fig. 6. Fuse wire system

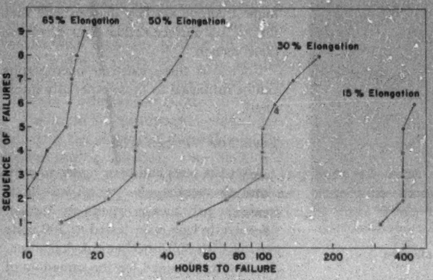


Fig. 7. Quantitative effects of electromechanical stress cracking; electric stress 200 volts/mil

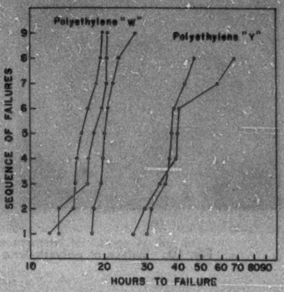


Fig. 8. Duplicability of test; measurements made with electric stress of 200 volts/mil and mechanical elongation of 50%

QUANTITATIVE EFFECTS OF MECHANICAL ELONGATION

To determine quantitatively the effect of different amounts of mechanical elongation of polyethylene resins, measurements were made on films under an electric stress of 200 volts per mil with 15, 30, 50, and 65% elongation. Fig. 7 shows that mechanical strain has a profound effect on the life of polyethylene exposed to a mild corona discharge. The samples with higher elongation had the shorter corona life. The corona energy in these experiments is low, the applied voltage being approximately 60 volts per mil above corona ignition. The experiment was repeated many times.

DUPLICABILITY OF TESTS

Fig. 8 is a plot of the life of each specimen shown in sequence of failure for five sets of samples of Du Pont polyethylene. There are three samples made from the same lot of polyethylene *W* and two samples from polyethylene *Y*. The failure times of the ten specimens of each of the three samples of *W* show little spread. There is a distinct difference demonstrated between resins *W* and *Y*. All these measurements were made under standard laboratory conditions at 200 volts per mil and 50% mechanical elongation.

Table I. Effect of Stress Annealing*

Material	Treatment	F (5/10), Hours
Polyethylene *P*	Tested as received	36.6
Polyethylene *P*	Stress relieved by annealing 6½ days at 70 C	Test stopped at 348 hours
Polyethylene *Q*	Tested as received	39.0
Polyethylene *Q*	Stress relieved as for polyethylene *P*	Test stopped at 324 hours

* Electrical stress 200 volts/mil, initial elongation 50%.

Table II. Effect of Air (50% RH) and Nitrogen (Dry) on Electromechanical Stress Cracking*

Material	F (5/10), Air, Hours	F (5/10), Nitrogen, Hours
Alathon 3, *NC-10*	36	113
Alathon 5, *NC-10*	35	140

* Room temperature; mechanical elongation 50%, electrical stress 200 volts/mil.

METHOD OF EXPRESSING RESULTS

Because of the great amount of numerical data accumulated, a convenient method was needed for compressing it. From this work and from experience on direct electrode studies on polymeric materials which have not been elongated, the time of the fifth failure out of ten specimens appears to be the most meaningful number to report. The expression F (5/10) hours is used to designate this quantity which is defined as the time in hours required for the fifth failure to take place when ten points on a sample are exposed simultaneously. Considerable time is saved by stopping the test after failure of the fifth specimen of the ten areas exposed.

EFFECT OF STRESS ANNEALING

The early measurements on the mandrel samples indicated that the life of samples was increased when some mechanical stress relaxation had taken place. A series of measurements were made on samples which were annealed after they had been mechanically stressed. The samples were mounted on the lower electrodes and clamped when elongated 50%. The electrodes and samples were placed in an oven for 150 hours at 70 C (degrees centigrade). The results are shown in Table I. The samples which were relaxed mechanically through this heat treatment exhibited a longer corona life.

A practical implication of these data is that the deleterious mechanical stresses in cables due to manufacturing operations or flexing during installation can be relieved by heat. A practical way to relieve these strains may be resistive heat loss in the conductor.

EFFECT OF AIR AND NITROGEN ON CORONA LIFE

Measurements of electromechanical stress cracking were made on many different polyethylene compositions at room

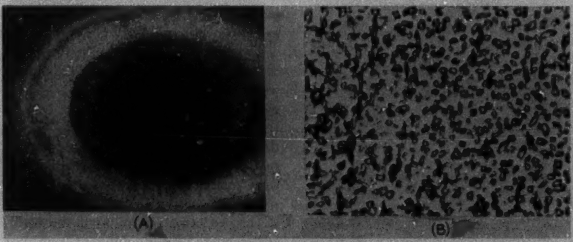


Fig. 9. Surface of polyethylene exposed to corona in air, standard laboratory conditions

A—5× magnification
B—200× magnification

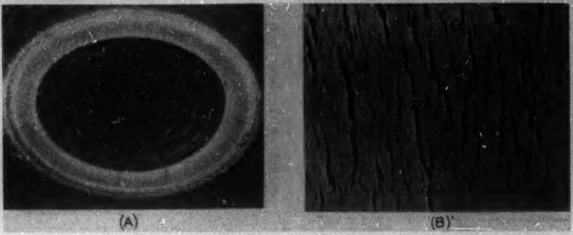


Fig. 10. Surface of polyethylene exposed to corona in dry nitrogen

A—5× magnification
B—200× magnification

temperature in both an air atmosphere (at approximately 50% RH) and a nitrogen environment at atmospheric pressure. Typical results are reported in Table II. The measurements made under an atmosphere of air always display a shorter corona life than those measured under nitrogen. In all tests which were run in air regardless of mechanical strain, a white ring was observed on the film around the outside edge of the electrode and the area subjected to the corona discharge. Many samples tested were examined under an optical microscope. Fig. 9(A) shows a typical sample (5× magnification) measured in air. The area affected by the corona appears as an oval because the film has been removed from the clamps of the lower electrode and has been allowed to relax. Fig. 9(B) is a view of the same sample but with a magnification of 200 diameters. The crystals were present on the surface of all films measured in air under standard laboratory conditions. Infrared analysis of these crystals indicated that they were oxalic acid. In addition to the crystals there are some cracks which always appear at right angle to the direction of elongation. It is of interest to note that S. I. Reynolds of the General Electric Company reported at the 1957 National Research Council Conference on Insulation that he had found oxalic acid crystals on the inside surfaces of voids in polyethylene which had been subjected to sufficient voltage to cause internal corona.

Fig. 10(A) is a view (magnified 5×) of the same material subjected to corona in a nitrogen atmosphere. This oval is smooth in appearance and does not show the crystals that are observed in the air sample. Fig. 10(B) is a 200-diameter magnification of the hazy area observed in Fig. 10(A). There are no crystals apparent. There are, however, many very fine surface cracks running transverse to the direction of mechanical stress. The mechanism of the failure under nitrogen is not yet completely understood.

Table III. Effect of Humidity on Electromechanical Stress Cracking, Alathon 3B Polyethylene Resin

Humidity, %	F (5510), Hours
20	18.0*
24	13.6
25	17.0
25	17.9
28	19.3
30	19.0
30	17.0
40	24.9
40	47.1*
54	36.5
54	36.5
54	39.3
60	45.9
60	48.0

* Sample cleaned with methyl alcohol before testing.

Note: Film thickness, 0.011 inch; electrode weight, 90.0 grams; mechanical elongation, 50%; electrical stress, 200 volts/mil.

ELECTRON MICROSCOPIC STUDY OF POLYETHYLENE EXPOSED TO CORONA IN DRY NITROGEN

An electron microscopic study was made on a polyethylene which was subjected to corona in a nitrogen environment at atmospheric pressure.

The results of this examination show that damage is confined to the area in the white ring formed around the periphery of the electrode. The area under the electrode was essentially unaffected. Within the white ring there were surface cracks running parallel to one another and at right angles to the direction of mechanical stress.

The mechanism of failure in dry nitrogen seems to be quite different from that in 50% RH air. As noted with optical microscopy, no oxalic acid crystals or other surface deposits are apparent. The surface cracks appearing in the region of highest corona activity demonstrate that failure is a surface rather than a volume phenomenon.

Copies of these electron micrographs can be furnished to those especially interested.

EFFECT OF WATER VAPOR

After the tests had been under way for about six months, an attempt was made to check early measurements on Alathon *3B* polyethylene resin tested at 200 volts per mil with 50% mechanical elongation. The results obtained for the same lots of film with all known variables controlled were most discouraging. The possibility of the sample slipping in the grip, variations in extrusions, and other factors which might explain the cause of the poor reproducibility were checked.

The only parameter not carefully controlled during this period was the relative humidity. Normally the tests are made under standard laboratory conditions, but during this period close control of the relative humidity was not possible because of construction work which allowed outside air to overload the air-conditioning system. A record of the relative humidity in the laboratory had been kept during the period the measurements were made. The data were correlated with humidity and tabulated in Table III. The only two measurements which do not fit are marked with asterisks. The records disclosed that these films had been cleaned with methyl alcohol before testing thus introducing another variable to the test.

Modifications were made to the test chamber so that moist or dry air could be maintained over the electrode areas. Scrubbing towers filled with water were used to wet the air which passed through the test chambers. The electrode systems were enclosed after water was poured on

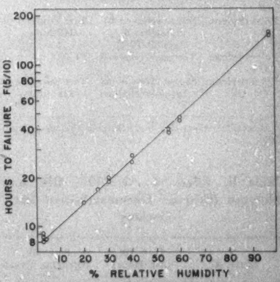


Fig. 11. Effect of humidity, measurements made with electric stress of 200 volts/mil and mechanical elongation of 50%

the floor of the box. A similar system was used to obtain dry air using Drierite in place of the water. Several measurements were made on Alathon *3B* in low (5%) and high (95%) RH. The results, along with the data obtained at the intermediate conditions, are plotted in Fig. 11.

An optical microscopic study was made of the surface of these films in the corona bombarded area. Figs. 12(A) and (B) show photomicrographs of the surface of the sample that failed in dry air. The sample failed by surface cracking, the cracks being at right angles to the direction of elongation. There is no evidence of oxalic acid. Figs. 13(A) and (B) show samples tested under a relative humidity of about 20%. Both cracks and oxalic acid crystals are observed. The oxalic acid crystals, however, are quite small and many liquid droplets appear. Figs. 13(C) and (D) are samples that were measured at 50% RH. Both cracks and oxalic acid are observed. The cracks are larger than those shown in the sample under corona at 20% RH.

The fifth failure for the sample which had been exposed to a wet-air (approximately 95% RH) atmosphere was after 155 hours. Figs. 13(E) and (F) are photomicrographs of the surface of the films. Minute cracks which were observed while the film was under tension are now hidden by wrinkles on the top surface. The surface resistivity of the area affected by the corona ranged from 10^6 (wet surface) to 10^{15} (dry surface) ohms per square compared to 10^{11} ohms per square for the area where there had been no corona.

The value of dissipation factor through the film in the area of corona bombardment was determined to be less than 0.0001 at 1 kc, indicating little or no volume effect.

EFFECT OF WET NITROGEN ON CORONA LIFE

Because such a great difference was found between wet and dry air, the effect of humidity on life in a nitrogen atmosphere was investigated by using nitrogen at 95% RH in place of air. A phenomenally long life was obtained for Alathon *3B* under a stress of 200 volts per mil. No failures had been recorded at 575 hours. To accelerate failure, the voltage was increased by 50%; four of the ten samples failed shortly and the remaining six did not fail in an additional 265 hours. The voltage was next raised to 400 volts per mil; two samples failed immediately, three within 40 hours, and the lone survivor had not failed before the test was discontinued after 120 hours at

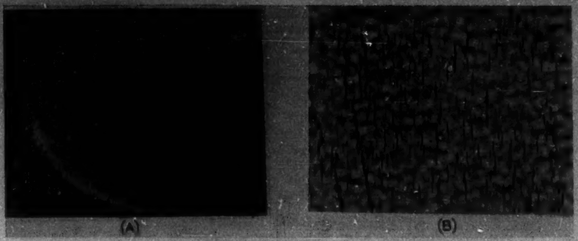


Fig. 12. Surface of polyethylene exposed to corona in dry air

A—5× magnification
B—350× magnification

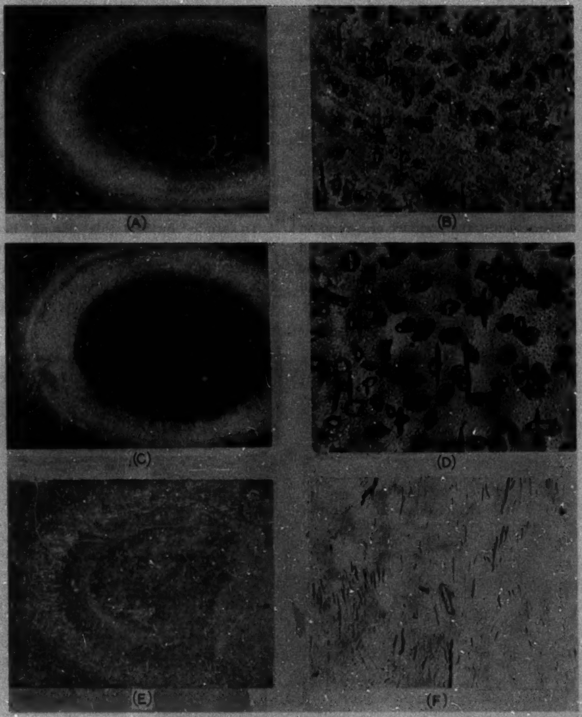


Fig. 13. Surface of polyethylene exposed to corona in air

A—5× magnification, 20% relative humidity
B—350× magnification, 20% relative humidity
C—5× magnification, 50% relative humidity
D—350× magnification, 50% relative humidity
E—5× magnification, 95% relative humidity
F—350× magnification, 95% relative humidity

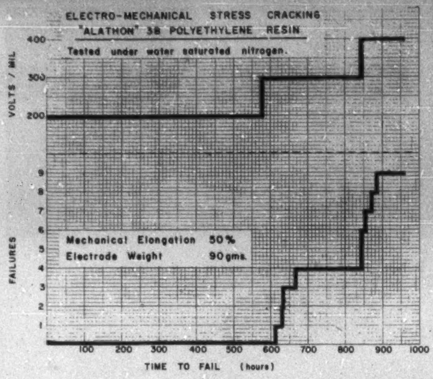


Fig. 14. Tests under nitrogen atmosphere

Table IV. Effect of Carbon Dioxide at Atmospheric Pressure on Electromechanical Stress Cracking, Alathon 3B Polyethylene Resin

	Carbon Dioxide	Air	Nitrogen
Dry	3.1	9	113
Wet	447	155	500

Mechanical elongation 50%, electrical stress 200 volts/mil.

this voltage. The results of the experiment are shown graphically in Fig. 14 which is a plot of the sequence of failures as a function of the hours to failure. The upper plot is a record of the applied voltages. These data are for individual test points.

The variation of intensity of the corona observed in a wet nitrogen atmosphere is very interesting. During the corona life measurements, the presence of corona on the sample was checked with a radio tuned to 1 mc (megacycle) and placed close to the test bay. One great difference existed between this test and previous tests. Shortly after the experiment was started, say about 1 hour, the radio ceased to indicate a noise caused by corona. No visual indication of the corona at the electrodes was observed. Periodically the radio did indicate corona bursts of several seconds duration and corona was visible at some test points during this time. This action occurred approximately every 5 minutes.

To study further this corona behavior in wet nitrogen, a new sample of Alathon *3B* was placed in the test chamber. A Biddle corona test set was used to measure corona starting voltage. The voltage was increased slowly. Corona appeared at approximately 140 volts per mil and after a few seconds at this voltage, disappeared. On raising the voltage level to 160 volts per mil corona again appeared but disappeared within a few seconds. This procedure was repeated until a steady corona discharge was finally obtained at 350 volts per mil.

These observations indicate a mechanism for longer life under wet conditions. The initial corona discharges produce a chemical reaction at the polymer surfaces to produce semiconducting films which "grade" the potentials in such a way as to eliminate the corona. After periodic corona bursts occur, the surface of the film is "healed" and the surface is given protection for a short period.

Electron Microscopic Study of Corona Damage in Moist Nitrogen

The surface of a polyethylene film after failure in the moist nitrogen atmosphere

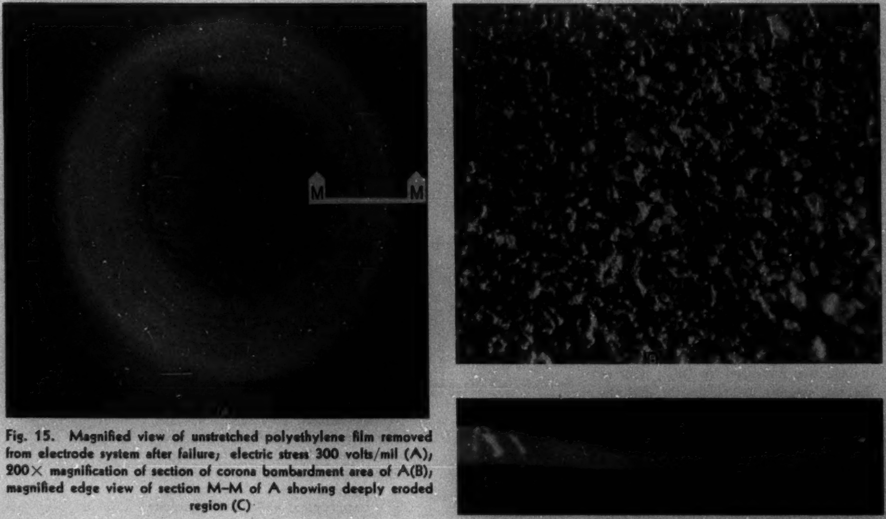


Fig. 15. Magnified view of unstretched polyethylene film removed from electrode system after failure; electric stress 300 volts/mil (A); 200× magnification of section of corona bombardment area of A(B); magnified edge view of section M–M of A showing deeply eroded region (C)

Fig. 16. Magnified edge view of sample exposed 600 hours at 200 volts/mil. Slight surface erosion may be detected by placing straight-edge along the top surface profile

Table V. Electromechanical Stress Cracking, Alathon 3B Polyethylene Resin

Electrical Stress, Volts/Mil	Exposure Period, Hours	Failures
400	504	None
600	504	None
800	500	None
1,000	500	None
1,200	10,000+	None

Note: Film thickness, 0.011 inch; electrode weight, 90.0 grams; mechanical elongation, 50%. Tested under Dow Corning Corporation 200 fluid (2 centistokes).

was examined in order to determine the reason for the improved corona life as compared with previous tests in dry nitrogen. Surface replicas at low (5X) magnification showed gross differences between the surface markings of the two samples. Selected areas were then examined by the electron microscopic technique to show details of the surface damage. A series of electron micrographs show that the damage to the surface of the sample measured in the wet nitrogen is more extensive and less well-defined than that to the sample measured in dry nitrogen. This fact and the observation of the characteristic tree-shaped markings support the belief that the improved test life is due to grading of potential by a semiconducting film.

Copies of these electron micrographs are available for those interested.

EFFECT OF CARBON DIOXIDE AT ATMOSPHERIC PRESSURE

The results of measurements of Alathon *3B* in carbon dioxide at atmospheric pressure are given in Table IV. The shorter life for the resin in dry carbon dioxide may be attributed to the higher corona intensity obtained in this gas than for nitrogen or air at the same voltage stress. The intermediate value obtained for wet carbon dioxide may be the result of the formation of carbonic acid on the surface of the film providing a protective semiconductive layer.

RESULTS WITH ZERO ELONGATION

Fig. 7, discussed earlier, shows the results of a series of experiments on the effect of mechanical elongation on life under corona. No data on failures were given for zero elongation since, at the 200-volts-per-mil level, sample testing was discontinued after 600 hours without any failures. Tests were begun on new samples at a stress of 300 volts per mil. The F(5/10) value (time of failure of the fifth sample of ten samples tested in parallel) at 300 volts per mil was 610 hours.

Microscopic examination of samples which failed under corona showed radically different surface conditions for polyethylene which was elongated than that which was not elongated. In the case of an elongated sample, the surface exposed to the corona discharge is covered with fine cracks which run at right angles to the direction of stretching. In the unstretched sample the area exposed to corona discharge is deeply eroded. It is believed this wearing away is caused by the impingement of gaseous ions formed in the corona discharge.

Fig. 15(A) is a slightly magnified view of the surface of an unstretched film removed from an electrode system after failure. As before, there was no corona and no change in the film under the central portion of the electrode. Failure occurred at the electrode edge.

Fig. 15(B) shows a highly magnified view of the corona bombarded area. Transverse cracks, observed with elongated samples, were not found. There are, however, oxalic acid crystals and pitted areas in the affected region. Oxalic acid crystals are more sharply defined near the edge of the bombarded area where the corona intensity was not as severe.

A section (M–M) was cut through the affected area of the sample shown in Fig. 15(A). In the magnified edge view of this section, shown in Fig. 15(C), it can be seen that the thickness in the area of most severe corona is reduced by 50% as a result of ion bombardment (erosion). The reduced thickness obviously leads to ultimate dielectric failure.

Examination of samples of the same lot of resin exposed for 600 hours to 200 volts per mil without failure showed this same effect of erosion but to a much smaller degree. A section cut radially from the corona bombarded area is shown in Fig. 16. Only a slight depression is apparent as compared to Fig. 15(C).

VOLTAGE ENDURANCE TESTING UNDER OIL

Electromechanical stress-crack measurements were made on Alathon *3B* under Dow Corning Corporation silicone fluid (dielectric constant 2.2) to study the effect of an electrical stress in the absence of corona. An initial voltage of 400 volts per mil was applied to the sample which was stretched 50%. The results of this experiment are shown in Table V. No failures were observed at 400 volts per mil after an exposure of 500 hours. The voltage was increased to 600 volts per mil, and every 500 hours thereafter an increase of an additional 200 volts per mil was made until 1,200 volts per mil (limit of the equipment) was obtained. Exposure time has exceeded 3,000 hours at 1,200 volts per mil. There have been no failures to date.

EFFECT OF ADDITION OF CARBON BLACK

Samples of Alathon *4* (polyethylene containing 2.5% carbon black) were tested at an electric stress of 200 volts per mil and under an elongation of 50%, and the results were compared to the base resin. The data show that the addition of carbon black reduces the corona life of the polyethylene to approximately one-half that of the clear resin.

Important Test Variables Which Must Be Controlled

In measuring electromechanical stress properties of polyethylene film, it has been found that careful control of the electrode weight, test voltage, humidity, and temperature are extremely important. All the samples must be of the same thickness and the surface must be clean.

EFFECT OF ELECTRODE WEIGHT

The upper electrode must make intimate contact with the film. If the electrode is too light, only a portion of its surface makes contact with the film and the area exposed to corona bombardment is not reproducible. In early experiments an electrode which had a total weight of 42 grams was used. When this weight found to be insufficient, a 90-gram electrode (1.59 square inches in area) was adopted and has been found to be quite satisfactory. The force is great enough to provide intimate contact without causing the electrode to dent the surface of the film.

EFFECT OF ELECTRODE METALS

All direct electrode measurements discussed in this study were made using electrodes of type *410* stainless steel. A series of experiments has been started to evaluate the effect on the corona life of polyethylene of electrodes fabricated from other metals.

Preliminary measurements in an atmosphere of wet air (95% RH) showed an increase of corona life from F (5/10) = 137 hours for the stainless steel electrodes to 217 hours for the copper electrodes. A heavy deposit of black copper oxide covered the latter electrode at the end of the test. The copper may catalyze the oxidation reaction on the surface of the polymer, thus assisting the formation of a protective semiconductive surface.

Need for Temperature and Humidity Control

The humidity of the atmosphere in the corona life test chamber must be carefully controlled for the reasons already indicated in detail. It is well to record the humidity and test temperature continuously. These tests have all been conducted at 23 C.

Sample Preparation

Careful sample preparation is very important. The width and thickness of the film should be the same for all samples tested. The rate of elongation must be low (1 inch per minute was used in these tests) and the amount of elongation must be accurately determined. A load cell and recorder are used with the stretching device to record the stress imparted to the film. Electrodes must be carefully cleaned and polished after each test to free them from pits or burrs.

Cleaning of Samples

The method of cleaning the sample before the test can affect the corona life. For example, cleaning polyethylene with methyl alcohol increases the life. Cleaning with distilled water will also increase life slightly. The samples used in the study were not cleaned, but were used just as formed by extrusion (except the two cases indicated in Table III). They were protected to keep the surfaces free from dust.

Projected Program

Despite the large amount of data accumulated and presented in this paper, there is much to be done to complete the study for polyethylene. In particular, methods of measuring corona energy density must be devised and this quantity related to corona resistance performance. Also the investigation will extend to other polymers.

References

1. Considerations in Specifying Corona Tests, **C. W. Ross, E. B. Curdts.** *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 75, Apr. 1956, pp. 63–67.

2. Effect of Electric Discharges on the Breakdown of Solid Insulation, **T. W. Dakin, H. M. Philofsky, W. C. Divens.** *Ibid.*, pt. I (*Communication and Electronics*), vol. 73, May 1954, pp. 155–62.

3. The Deterioration and Breakdown of Dielectrics Resulting from Internal Discharges, **J. H. Mason.** *Journal*, Institution of Electrical Engineers, London, England, vol. 98, pt. I, 1951, pp. 44–59.

4. Electrical Discharge Characteristics of the Plane Parallel Gaps Bounded by Insulating Surfaces, **J. C. Devins.** *AIEE CP57-83* (available on request).

5. How Humidity Affects Insulation, Part I, **R. F. Field.** *General Radio Experimenter*, Cambridge, Mass., vol. 20, 1945, pp. 6–12.

6. Corona Properties of Insulating Materials, **W. T. Starr.** *Electrical Manufacturing*, New York, N. Y., vol. 57, no. 6, June 1956, pp. 132–39.

7. Internal Discharges in Dielectrics: Their Observation and Analysis, **A. E. W. Austen, W. Hackett.** *Journal*, Institution of Electrical Engineers, vol. 91, 1944, pp. 298–322.

The Design and Manufacture of Direct Burial Wire

J. L. ROBB
ASSOCIATE MEMBER AIEE

W. L. ROBERTS
NONMEMBER AIEE

OUTSIDE the high-density service areas in the central parts of cities, telephone operating companies in this country have, until recent years, found it economically necessary to use aerial distribution facilities. Open-wire construction in particular, with its high initial cost, high maintenance cost, and frequent service interruptions, offered a desirable area for development of a direct burial facility. Various efforts using rubber insulated wire with no outer protection failed, which led to the development of rubber insulated wires with neoprene jackets. Experience in certain areas indicated the need for additional mechanical and electrical protection so a light, flat ribbon steel armor is provided under the jacket in certain types of this wire.

No. *16* Awg (American wire gage) copper conductors are used in these wires and this factor, coupled with the costs of the rubber and neoprene, results in a fairly costly product which economically prevents its use in many situations. In view of this condition, the development of lower cost facility was started in 1956.

Performance Requirements

Engineers of the Rural Electrification Administration (REA) anticipated extensive use of a low-cost buried-wire facility if such could be developed and therefore established performance requirements for a wire which would meet the needs of the REA telephone program:

1. Attenuation and d-c resistance must be small enough to permit total loop lengths, including some cable plant, to reach a large proportion of subscribers, who would normally be served by 1- or 2-circuit open-wire construction, without exceeding transmission limits and supervisory limits for presently available central-office equipment. Loops of about 15 miles in length would appear to be needed.
2. All materials used in the wire should be suitable for use both above and below ground, and a life expectancy of 20 to 25 years for the wire would be necessary.
3. The wire must be armored for protection against small rodents and termites.
4. A shield for lightning protection and noise mitigation must be used and a tentative shield resistance of 15 ohms per 1,000 feet was established.
5. The wire must have a breaking strength of approximately 500 pounds to permit installation with high-speed wire-plowing equipment and to withstand ground heaving and frost cracks. A further consideration was the intended use of this wire for "aerial inserts" and for aerial drop wire.
6. The wire will be in contact with treated wood poles at terminal points and must withstand exposure to pole-treating chemi-

---

Paper 59-777, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 24, 1959; made available for printing April 9, 1959.

J. L. Robb and W. L. Roberts are with Superior Cable Corporation, Hickory, N. C.

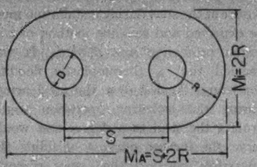


Fig. 1. Cross section of flat oval circuit

cals, particularly creosote and pentachlorophenol.

7. The cost of the wire in place must be competitive with first costs for open-wire construction.

8. Some carrier circuit potential would be desirable, but not mandatory.

A study by REA engineers indicated that a wire with no. *19* Awg copper conductors and a mutual capacitance of 0.083 μf (microfarad) per mile would meet transmission and supervisory requirements. By coincidence these are the characteristics of a no. 19 Awg cable pair. This same study proposed the use of black weather-resistant polyethylene as the insulant and offered suggested dimensions for the insulated core so proportioned that the minimum amount of insulating material would be used to achieve the desired mutual capacitance. The study also proposed the use of flat ribbon steel as a combined armor-shield-strength member.

Design of Buried Wire

Economy dictated that the two conductors in this wire be insulated in one operation, therefore a design using a parallel pair of conductors with black polyethylene extruded over them was mandatory.

SHAPE AND DIMENSIONS OF CORE

A flat oval core appeared to offer the most economic design from the standpoint of minimum amount of insulant needed to achieve the desired mutual capacitance, but since a flat steel-ribbon armor must be applied over this core a more nearly round core had to be used. Formulas for proportioning of a round shielded pair to achieve a desired mutual capacitance have been published but there was no information available on the proportioning of a core which was approximately elliptical.

Greene, Leibe, and Curtis, in an article appearing in the *Bell System Technical Journal*[1] have indicated that to obtain minimum capacitance for a flat oval shielded pair a spacing ratio of approximately 0.47 should be used. The spacing ratio is defined, in the case of the flat oval (Fig. 1), as the quotient obtained by dividing the interaxial conductor spacing (S) by the major diameter (M_a). The mutual capacitance of the flat oval circuit is very slightly greater than that of the shielded double coaxial circuit with an equal cross-sectional area. The total cross section of the flat oval circuit is

$$A=2RS+\pi R^2 \qquad (1)$$

where

$$M_a=S+2R$$

One conductor of a double coaxial circuit of equal cross section would have a radius of

$$R_1=\sqrt{\frac{1}{2\pi}(2RS+\pi R^2)}=R\sqrt{\frac{S}{\pi R}+\frac{1}{2}} \qquad (2)$$

The capacitance of the double coaxial circuit is given by the well-known relation

$$C_1=\frac{0.0194\epsilon}{\log_{10}\frac{2R_1}{d}}\ \mu\text{f per mile}$$

where d is the conductor diameter.

It is obvious from Fig. 1 that the capacitance of the flat oval circuit is less than that of a double coaxial circuit with similar insulation whose outside radius is $S/2$ (where $S<2R$). Then, if C represents the capacitance of the flat oval circuit,

$$\frac{0.0194\epsilon}{\log_{10}\frac{2R}{d}\sqrt{\frac{S}{\pi R}+\frac{1}{2}}}<C<\frac{0.0194\epsilon}{\log_{10}\frac{S}{d}} \qquad (3)$$

Furthermore, if 0.47 represents the optimum spacing for minimum capacitance, a similar oval of equal cross section, but having a spacing ratio of 0.50, would have a greater capacitance. Then

$$C<C'<\frac{0.0194\epsilon}{\log_{10}\frac{S'}{d}}$$

where

$$S'=2R'$$

Since the cross sections are equal, $2R'S'+\pi R'^2=2RS+\pi R^2$. After substituting $S'=(1.091)S$, equation 3 may then be reduced to

$$\frac{0.0194\epsilon}{\log_{10}1.165\frac{S}{d}}<C<\frac{0.0194\epsilon}{\log_{10}1.091\frac{S}{d}} \qquad (4)$$

where

$$\frac{S}{S+2R}=0.47$$

For a mutual capacitance of 0.090 μf per mile or smaller and an insulation with a dielectric constant of 2.5 or greater, the average value of those obtained from equation 4 must be correct to within ±3.5%.

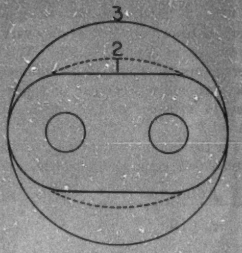


Fig. 2. Comparison of shield configurations

1—Flat oval
2—Quasi-elliptical
3—Circular

As indicated previously, the flat oval configuration presents manufacturing difficulties in the armoring operation. The quasi-elliptical cross section used may be regarded as a compromise between the flat oval circuit and the circular-shield balanced pair.

The capacitance of a balanced pair with a circular shield whose diameter is equal to the major diameter of the flat oval (Fig. 2) is given approximately by:

$$C_2\approx\frac{0.0194\epsilon}{\log_{10}\frac{2S}{d}\left[\frac{\sigma^2-1}{\sigma^2+1}\right]}$$

where

$$\frac{D}{S}=\sigma=\frac{S+2R}{S}$$

This reduces to

$$C\approx\frac{0.0194\epsilon}{\log_{10}1.276\frac{S}{d}} \qquad (5)$$

when

$$\frac{S}{S+2R}=0.47$$

For shield configurations lying within the area bounded by the flat oval and circular shields, the capacitance of the circuit is less than the upper limit indicated by equation 4, and greater than the value obtained from equation 5.

The following formula has been found to be accurate to within ±2.3% when applied to constructions with M_i/M_a ratios ranging from 0.52 to 0.75, and spacing ratios between 0.45 and 0.51:

$$C\approx\frac{0.0194\epsilon}{\log_{10}\frac{2S}{d}\left[\frac{M_a^2-S^2}{M_a^2+S^2}\right]-0.167\log_{10}\frac{M_a}{M_i}}\ \mu\text{f per mile} \qquad (6)$$

(The dielectric constant of general purpose black polyethylene may vary by as much or more than ± 2.3%. This deviation appears to be due to variation in the carbon-black content and perhaps also to moisture content of the compound.)

For the flat oval circuit with a spacing ratio of 0.47, equation 6 can be reduced to

$$C \approx \frac{0.0194\epsilon}{\log_{10} 1.146\frac{S}{d}} \text{ μf per mile} \qquad (7)$$

Equation 7 appears consistent with the inequalities expressed in equation 4. Equation 6 does not account for the shift of the optimum spacing ratio from 0.486 to 0.47, as the shield configuration changes from a circle to a flat oval.

Fig. 3 shows calculated mutual capacitance as a function of S/d for several M_t/M_a ratios with the spacing ratio fixed at 0.47.

Capacitance and Resistance Unbalance

When buried wire is used in the field, it must be anticipated that open-wire circuits and drop wires will be connected to the buried wire and therefore noise voltages induced in these facilities will be impressed on the buried wire. To minimize the effects of these noise voltages, capacitance unbalance to the shield and conductor resistance unbalance must be held to low levels.

In the previously mentioned rubber-neoprene buried wire, a capacitance unbalance maximum of 500 μμf per 1,000 feet is permitted. This seemed to be a reasonable figure and was adopted for the polyethylene insulated wire. In the case of the rubber-neoprene wire, however, this value is met by measuring the capacitance unbalance on a completed reel of wire and then, if the maximum value is exceeded, cutting the wire at the midpoint and splicing it together again with the conductors reversed. It was felt that this procedure was not necessary with the polyethylene insulated wire since it was anticipated that precise control of the polyethylene extrusion would produce wire within the capacitance unbalance limits. Manufacturing experience, however, indicated that a higher maximum value was needed to avoid excessive rejected material so a maximum value of 750 μμf per 1,000 feet was adopted.

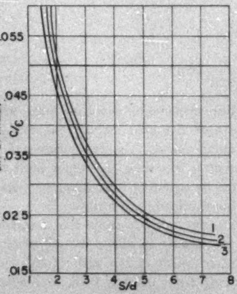


Fig. 3. Mutual capacitance graphs

1—Flat oval circuit
2—Quasi-elliptical circuit
3—Circular shielded circuit (spacing ratio = 0.47; cross sections not equal)

When this higher value of maximum capacitance unbalance was adopted, installation procedures in the field were changed to require that the two conductors of the buried wire be reversed at every terminal point. Since terminals appear, on the average, every 1/4 mile, this should control both capacitance unbalance and resistance unbalance very adequately.

Since a reversal of conductors at every terminal is required, it is necessary to identify each conductor in the buried wire. An easy and convenient means of doing this is to have one conductor tinned or plated.

The resistance unbalance limit for this wire is based on experience with other no. *19* Awg copper conductors and is set at 1.5 ohms maximum per 1,000 feet.

Armor-Shield-Strength Member

The requirements for protection against rodents and termites, along with the breaking strength requirement for the buried wire, pointed to the use of steel for this function since low cost was of such primary importance.

Based on measurements made on samples produced early in the development program, using steel armor, a shield resistance of 15 ohms per 1,000 feet appeared to be obtainable but later large scale production showed that the steel chemistry needed to meet the electrical and tensile properties was quite precise, so rather than sacrifice the electrical properties of the shield, which are at best marginal, the breaking strength requirement was reduced to 400 pounds.

Efforts to improve the conductivity of the shield and at the same time provide a measure of corrosion resistance by the use of galvanized or copper-covered steel were unsuccessful due to the cost of these materials.

It was recognized, however, that damage to the jacket of the wire and exposure of the steel during installation must be expected and so some method of corrosion protection was needed. It was decided that a thermoplastic flooding compound flooded over the steel armor during the jacketing operation would protect the steel to a degree and would serve the further purpose of filling the interstices between the armor ribbons, thereby preventing the entrance of water under the jacket. The flooding compound provided a very low-cost answer to this problem, but created another problem.

One intended use of this wire was as "aerial drop wire" in buried plant construction. Drop wires are attached to buildings and should be able to withstand a vertical flame test. The presence of the flooding compound causes the wire to be nonself-extinguishing. A type of flooding compound containing a flame retardant is currently under development which, it is believed, will overcome this difficulty.

Buried Wire Jacket

Because of the aforementioned self-extinguishing need, a polyvinyl-chloride jacket was used (see Fig. 4) This jacket provides the necessary toughness and will stand reasonable exposure to the common pole-treating chemicals, as well as provide the desired life above and below the ground.

The jacket offers no resistance to rodents or termites. The effects of soil bacteria and fungi on these jackets have not been evaluated, but no difficulties are anticipated from these sources.

Now that this wire has been in use in the field, it has developed that a method of indicating the footage on the surface of the jacket would be desirable. Since the route of the buried wire during installation is somewhat devious, conventional methods of chaining the line to determine the amount of wire used are inaccurate. Polyvinyl chloride lends itself to this type of surface marking and preliminary developments indicate that it

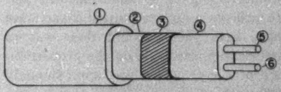


Fig. 4. Direct burial wire

1—Polyvinyl-chloride jacket
2—Thermoplastic flooding compound
3—Flat steel armor shield
4—Polyethylene core
5—Bare copper conductor
6—Tinned copper conductor

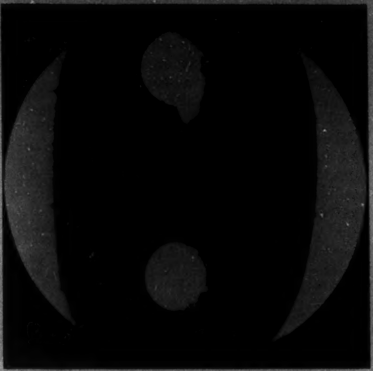


Fig. 5. Core cross section showing void formations adjacent to conductor surfaces

will be feasible to provide this marking on the wire.

Manufacture of Buried Wire

The manufacturing techniques used to produce buried wire are standard extrusion and wire-armor-serving techniques, but there are certain problems which have occurred that merit discussion.

CONDUCTORS

The application of tin or other materials as a plating on one of the copper conductors for polarity identification in this wire causes certain problems.

It is standard practice to draw wire to the finished size and then to apply the tin over this diameter. As a result, tinned conductors will be very slightly larger in diameter than bare wire of the same gage number. Also, standard specifications for tinned wire in no. *19* Awg permit a plus tolerance of 3%, while for bare wire the plus tolerance is 1%.

As a result of these factors, the tinned wire runs slightly larger than the bare wire and when these two different wires are paired, the direct capacitance to the shield is consistently higher for the tinned conductor.

This built-in unbalance, when added to the normal variations that occur in the extrusion of the core and application of the armor, results in unbalances which are greater than the permissible maximum for an undesirable percentage of the total wire produced. To overcome this, it is necessary that the tinned conductor be the same diameter as the bare conductor within close tolerances.

When two conductors of the same diameter are used, the capacitance unbalances become random rather than systematic, and the values are within limits.

Reducing the diameter of the tinned conductor, however, causes this conductor to run consistently high in resistance but still within resistance unbalance limits. This situation is, of course, corrected by the practice of reversing conductors at terminal points in the field.

EXTRUSION OF THE CORE

Considerable difficulty was experienced in verifying the calculated design for the desired mutual capacitance due to presence of voids along the conductor surface (see Fig. 5). These voids along the conductor surface tended to lower the mutual capacitance significantly. The presence of the voids in the insulation lowers the dielectric strength and cannot be controlled to yield consistent values of mutual capacitance, so it was necessary to eliminate this condition.

The effect of the voids on mutual capacitance was calculated using a simple coaxial structure with a composite dielectric since it was felt that this would be indicative of the order of magnitude of mutual capacitance variation that could be expected with the 2-conductor shielded pair.

For purposes of comparison, a coaxial type construction with a no. *16* Awg conductor and 0.100 inch wall of polyethylene with $\epsilon=2.6$ was chosen.

The potential difference formula as given in reference 2, p. 80, is:

$$V_1-V_2=\frac{\rho_L}{2}\left(\frac{1}{\epsilon_b}\ln\frac{r_2}{r_1}+\frac{1}{\epsilon_c}\ln\frac{r_3}{r_2}\right)$$

This yields a capacitance of:

$$C=\frac{0.0388}{\frac{1}{\epsilon_b}\log\frac{r_2}{r_1}+\frac{1}{\epsilon_c}\log\frac{r_3}{r_2}}\ \mu\text{f per mile}$$

where

ϵ_b = dielectric constant of void = 1
ϵ_c = dielectric constant of polyethylene, 2.6
r_1 = conductor radius, or diameter
r_2 = dielectric boundary radius, or diameter
r_3 = shield radius, or diameter

The construction cited no. *16* Awg, 0.100-inch wall, $\epsilon_c=2.6$) yields the following values:

For no void space around conductor: $C=0.144$ μf per mile.
For 0.001-inch average void space around conductor: $C=0.139$ μf per mile.
For 0.002-inch average void space around conductor: $C=0.134$ μf per mile.
For 0.003-inch average void space around conductor: $C=0.129$ μf per mile.

The capacitance decreases expressed as percentages amount to:

For 0.001-inch void space: 3.5%.
For 0.002-inch void space: 6.9%.
For 0.003-inch void space: 10.4%.

The presence of the voids seemed to be due to the presence of moisture in the carbon black used in the black polyethylene so a series of experimental runs were conducted starting with natural polyethylene and proceeding with black compounds having 0.025%, 0.041%, 0.058%, 0.075%, and 0.088% moisture contents, respectively. The natural polyethylene yielded a smooth, dense, void-free core as did the black polyethylene with a moisture content of 0.025%. The compound with 0.041% moisture content was relatively free of voids but all the other compounds resulted in gross void formation.

In those compounds where voids formed, it was found that high conductor preheat temperatures aggravated the void situation while removing the preheat caused the insulation to shrink away from the conductor creating an equally unfavorable condition. With the low moisture content compounds, the upper preheat temperature was not limited by the appearance of voids.

To control the void problem it is necessary to take steps to ensure that as little moisture as possible be absorbed by the compound prior to its use and to use extrusion techniques to remove as much moisture as possible during extrusion. Hopper dryers, fine mesh screen packs to improve mixing and increase back pressure, and tools designed to increase head pressure are all beneficial.

Conclusion

Over 1,000 miles of this type of buried wire have been installed in the field and experience to date indicates that the design objectives of this wire were sound. The carrier capabilities of this facility are quite limited but electronic developments may eliminate this shortcoming. Such developments may also extend the loop-limit capabilities of this wire.

The design of buried wire has stabilized and manufacturing problems have been solved so that this product is now available as a standard outside plant facility at a cost which compares favorably with open-wire construction in those areas where high-speed wire plows can be used. It is reasonable to expect that in coming years a substantial amount of this type of telephone distribution facility will be buried directly in the ground.

References

1. The Proportioning of Shielded Circuits for Minimum High-Frequency Attenuation, E. I. Green, F. A. Leibe, H. E. Curtis. *Bell System Technical Journal*, New York, N. Y., vol. 15, no. 2, Apr. 1936, pp. 248–83.

2. Electric and Magnetic Fields (book), S. S. Attwood. John Wiley & Sons, Inc., New York, N. Y., 1949, pp. 79–80.

3. Buried Telephone Plant in Residential Areas, W. J. Lally, C. C. Lawson. *Bell Laboratories Record*, New York, N. Y., vol. 37, no. 3, Mar. 1959, pp. 105–09.

4. Design of Wire and Cable for Direct Burial in Rural Telephone Systems of REA Borrowers, R. S. Baily, C. R. Ballard, F. F. Farnsworth, A. L. Richey. Sixth Annual Symposium, U. S. Signal Corps, Washington, D. C., 1957.

5. Communication Underground Distribution Plant in Urban Areas, C. H. Elder. *Ibid.*

6. Buried Rural Telephone Distribution Systems, D. G. Wilson. *Ibid.*

7. Buried Cable Telephone Systems, G. L. Chilberg. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, May 1957, pp. 130–35.

8. The All-Buried Telephone Service Concept, J. S. Reed. *Telephone Engineer and Management*, Chicago, Ill., vol. 62, nos. 7 and 8, Apr. 15, May 1, 1958.

◆

Discussion

R. W. Perkinpine (American Telephone and Telegraph Company, Kansas City, Mo.): The proposed buried wire may meet a need for a relatively inexpensive wire system which could be used for rural telephone distribution purposes. In order to meet the cost objectives, sacrifices in transmission and breaking strength have had to be accepted.

In sections of the country where rodents, particularly gophers, are present this wire might not be satisfactory. Where rodent damage is a possibility, it might be preferable to consider other more costly buried wire now available. However, from the facts and figures covered in the paper, it is felt that this wire would amply meet the performance and other requirements of the REA telephone program in many sections of the country.

All-Transistor Magnetic-Core Memories

BEN T. GODA
NONMEMBER AIEE

WILLIAM ROBERT JOHNSTON
NONMEMBER AIEE

SEYMOUR MARKOWITZ
NONMEMBER AIEE

MILTON ROSENBERG
NONMEMBER AIEE

RAYMOND STUART-WILLIAMS
NONMEMBER AIEE

RECENTLY WORK has been conducted on a medium-capacity high-speed magnetic-core memory, employing small ferrite cores as the storage element. In order to perform this work, it was necessary to conduct extensive tests on the behavior of ferrite cores at high repetition rates. The experimental results permit the establishment of criteria for use of linear or coincident-current selection techniques.

The properties of special small ferrite cores which were developed for use in high-speed transistor-drive memories are discussed in this paper. The concomitant problems arising in the execution of an all-transistor approach are also considered.

Requirements for High-Speed Memories

The advent of high-speed transistor logic circuits which operate at clock rates of 1 to 10 mc (megacycles) has created a requirement for high-speed transistor-driven random-access storage units. Present commercial units operate in the 10- to 12-μsec (microsecond) range and special memories have been built operating as fast as 5 μsec per cycle. Few memories have been built which are transistor driven. Drive circuits must handle relatively high power, and for this reason it is much more complicated to design high-speed storage units than it is to design high-speed logic circuits.

There is an immediate requirement for a relatively small storage unit operating at the maximum possible speeds of transistors and magnetic cores. Secondary to this is a requirement for large memories operating at cycles faster than present memories. These will provide the bulk of the storage in a computer design.

It appears feasible to operate a fast, low-capacity memory with a 2- to 3-μsec cycle time, an intermediate capacity device with a 4-μsec cycle time, and a large memory with a 6- to 8-μsec cycle time. These cycle times are not as short as computer designers could use, but are as fast as is practicable with present transistors and magnetic cores.

AVAILABLE MATERIALS

Two basic classes of ferrite core material are presently available. The first is the rectangular loop material which is used in coincident-current core memories. This is commercially available from 1-μsec switching time to 5 μsec, and experimental quantities have been manufactured over a range from 1/4 μsec to 10 μsec. The rectangular loop materials can also be employed in linear systems and have been experimentally operated under these conditions at switching times as low as one-twentieth of a μsec. The second class of materials with less rectangular loops has been developed for use in linear systems. Much lower drive powers are required for a given switching time than is possible in rectangular materials.

MODES OF OPERATION

Two basic modes of operation are possible in high-speed storage devices. The first is the conventional coincident-current mode, while the second is the linear selec-

---

Paper 59-791, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted September 24, 1958; made available for printing April 15, 1959.

Ben T. Goda, William Robert Johnston, Seymour Markowitz, Milton Rosenberg, and Raymond Stuart-Williams are with Telemeter Magnetics, Inc., Los Angeles, Calif.

tion mode. Both modes may employ direct drive, diode switching, or magnetic switching. Transistors have altered the economics of memory construction in that they represent a relatively inexpensive way of providing direct drive at magnetic-core impedance levels to storage arrays. Therefore, direct drive may be used in many applications where switch drive was the only feasible approach in the past.

MEMORY CONFIGURATION

A typical coincident-current memory is 3-dimensional in nature. The address is decoded into an x-address and a y-address. The parallel bits representing the words represent the z axis. The capacity of the storage unit is xyz. The equipment in the unit is proportional to a function which can be expressed approximately as $A(x+y)+Bz$. The factor B is usually much larger than A. A linear selection memory is 2 dimensional in nature. The address is represented by one address (s). The capacity of the equipment is sz. The equipment in the unit is approximately a function of $Cs+Dz$, where D is usually larger than C.

In large systems the cubic nature of coincident-current operation yields the most economical system. In small or fast systems the linear selection mode can become more economical. This is particularly true in the case of long words, as D is usually smaller than B.

SPEED LIMITATIONS

Most high-speed core memories employ a basic 2-beat cycle. In the first beat, reading takes place and the core is set into the ZERO state. In the second beat, a writing action is performed and the core can be left in the ZERO state, or set into the ONE state.

There are three classes of speed limitations in a core memory. They are as follows:

1. Time taken to perform red tape and to propagate signals.
2. Time taken for actual magnetic-core operation.
3. Limitations due to heating of the cores and transistors at high operating rates.

LIMITATIONS DUE TO PROPAGATION

A large number of short time delays occur in any memory cycle. These are:

1. A new address is transmitted to the memory. It may be necessary to set an address register and it is always necessary to perform at least partial decoding of the address. This complete operation usually requires between 1/2 and 1 μsec.
2. The drivers are operated and these have finite rise time. Because of the very high currents employed, it is difficult to achieve rise times of less than 0.1 to 0.2 μsec.
3. If the memory is switch driven there is a propagation time in the switch itself. In diode switches this can be as low as 0.1 μsec, but in magnetic switches it is rarely lower than 0.2 μsec and can be as high as 0.5 μsec.
4. The drive currents are now applied to the memory stack and the first information output will appear some fixed time later than this. In a coincident-current memory the currents must propagate through either $x.z$ or $y.z$ cores and this introduces a delay of from 0.1 to 0.3 μsec in the average memory. In a linear selection memory there is virtually no propagation time as the read current passes through only z cores.
5. When the information signals are emitted from the matrices they must be amplified, gated, and then used to set a memory register. This usually takes between 0.3 and 0.5 μsec.

   In a coincident-current memory the drive pulse is so long that this delay is usually masked by the length of a drive pulse. In a linear selection memory very short pulses are employed and part of this delay must occur as extra cycle time.
6. The drive pulse is now cut off and there is a finite decay time. With correctly designed transistor circuits the fall time of the pulse may be between 0.2 and 0.3 μsec.
7. It is now possible to apply the digit or inhibit drive. The digit driver has a rise time of 0.1 to 0.2 μsec and this signal must propagate through $x.y$ cores in coincident-current operation, or s cores in linear selection operation. This time can vary from 0.1 to 0.5 μsec, depending on the size of the memory.
8. The write drive pulse is subject to the same delays—1, 2, and 3—as the read drive pulse. It cannot be commenced until the read drive pulse is completely finished and must not have achieved full current until delays 7 are complete.
9. There is then the period for the write drive pulse. This pulse and the digit drive pulse then decay in 0.2 to 0.5 μsec and the cycle is complete.

In a well designed equipment using available transistors it is difficult to reduce these propagation times to less than a total of about $1^1/_2$ μsec. Even in the next few years it is difficult to imagine this number being below 3/4 μsec.

TIME FOR MAGNETICS

Present 50-mil 1-μsec cores require $1^1/_2$-μsec pulses for both read and write actions. Therefore, the magnetics require 3 μsec for adequate operation. In a coincident-current cycle some overlapping of propagation time and magnetic time is possible and a 4-μsec cycle can just be achieved in small memories.

The 50-mil 1/2-μsec linear selection cores require 1/2 μsec for read and 1 μsec for write. Therefore, it is possible to achieve a 3-μsec cycle using linear selection. In very small memories it is just possible to overlap times to achieve a 2-μsec cycle.

These facts define one fundamental change-over point between coincident-current and linear selection systems. Below 4 μsec in small systems, and 5 or 6 μsec in large systems, linear selection must, at present, be employed regardless of cost.

LIMITATIONS DUE TO HEATING

Speed limitations of transistors due to overheating at high rates of operation is a well-known problem. Similar problems occur in the storage cores, but the problem is more complex in nature.

The ferrite storage cores are always strung upon a network of wires. These wires must be operated at low current density to ensure that they do not introduce heat into the system, but rather provide a conductive cooling medium for the storage cores. Ferrite cores are very poor electrical and thermal conductors, and heat dissipation within the core during an operating cycle is removed very slowly. In most present cores the operating cycle can be considered as a few μsec, whereas the thermal time constant is a few tens of milliseconds.

There are two manifestations of heating problems in ferrite core memories. First, in memories of all types, the storage properties are degraded as the temperature increases. The disturbed ONE becomes smaller and the disturbed ZERO and the nth disturb becomes larger as temperature increases. Ultimately the two types of signal become so close to each other that the memory cannot operate reliably. For an individual core, this point usually occurs somewhere in the region between 150 F (degrees Fahrenheit) and 250 F. The precise temperature depends on the material employed and is usually higher in the case of high-speed materials than in low-speed materials.

The second effect is also due to the change in storage properties as the temperature is changed. This effect occurs principally in coincident-current memories. In such memories an adequate signal to noise ratio can be obtained only if all cores in a plane have very uniform properties. Very slight local changes of temperature can degrade the performance of the plane. This effect is well known when considering the cooling of a magnetic system. Air blowing on part of a plane, and not all of it, will usually make that plane inoperative. It is not as generally realized that the same effect can occur at very high speeds of operation when a portion of a plane is scanned continuously.

This is the typical mode of operation in any random access memory used in conjunction with a computer. The result is that local heating can produce a very poor signal-noise ratio in the plane In large planes a few degrees Fahrenheit change in temperature locally can degrade performance This effect is increase as the over-all temperature of the plane is increased, thus reducing the over-all working margins.

In order to reduce the reliability of operation of a plane, it is not necessary to heat cores up to a very high temperature. A few degrees rise in a local area will be sufficiently dangerous. This effect is strictly a function of speed and program, and can be missed by many of the tests that are performed on random access memory units. The only formal test that will detect this effect is to scan a small portion of the memory very rapidly and to sample, rather rarely, cell outside that scanned area. When a failure occurs it usually appears as if the cell outside the scanned area has failed.

Some idea of the seriousness of this problem can be appreciated from the experimental fact that the normal 80-mil coincident-current cores cannot be reliably operated more rapidly than about 1 cycle every 3.6 μsec, even under the most closely controlled conditions of current and ambient temperature. The 50-mil core can just operate at 2 μsec per cycle.

Reduction of Operating Cycle

The minimum operating cycle time may be reduced by any or all of the following changes:

1. *Reduce the Size of the Core.* This increases the surface-to-volume ratio and hence improves the cooling of the core. In many transistor applications a 50-mil core is being used in place of the usual 80-mil core. Theoretically this core can operate satisfactorily under coincident-current conditions even when cycled as quickly as 3 μsec per cycle. Because of propagation limitations this core has not been operated under coincident-current conditions faster than 4 μsec per cycle.

2. *Employ a Special Mode of Operation.* In linear selection an asymetric loop is used instead of the symetric loop used in coincident-current operation, and this makes it possible to shorten the cycle time. The exact amount by which the cycle time can be shortened depends on many parameters.

3. *Improve the Thermal Properties of the Material.* This is a difficult problem but continuous progress is being made. Fortunately, the materials which show the most promise of improvement are the high-speed materials.

Cores Which Are Available for Transistor Drive

In limited applications the present 80-mil core can be transistor driven. A "low-noise" core is usually selected. This is a core which is very rectangular and which has very low *n*th disturb outputs. This minimizes the inductance that must be driven by the transistors. In high-speed applications 80-mil cores can be driven only if the memory capacity is small, or if special transistors in relatively short supply are used in the drive circuits. Small cores are much more suitable for transistor-driven systems.

The 50-mil core is 0.050 inch OD (outside diameter), 0.030 inch ID (inside diameter), and 0.015 inch thick. It is commercially available in three different materials.

ONE-μSEC COINCIDENT-CURRENT MATERIAL

The core is driven with approximately 500 ma (milliamperes) full current or 250 ma half current. Under these conditions the "knee" is above 300 ma. The core switches in almost exactly one μsec. The disturbed ONE output is about 65 millivolts and the disturbed ZERO output is about 9 millivolts. The *n*th disturb is about 1.6 mv (millivolts). These results are obtained with a current pulse which has a 0.2-μsec rise, 1.5-μsec top, and 0.2-μsec fall.

HALF-μSEC LINEAR-SELECTION MATERIAL

The core is driven with approximately 400 ma full current for read and 100 ma half current for write. The core switches during reading in about 1/2 μsec and during writing in about 1 μsec. The ONE output during reading is 75 mv and the ZERO output is 15 mv. The write outputs are about 30 mv and the half disturb output associated with the digit drive operation is about 2 mv per core. The read pulse must rise in 0.1 μsec, remain up 0.4 μsec, and fall in 0.1 μsec. The write pulse must have a top that is a little over 1 μsec in duration. If a slow rise and fall time can be permitted during writing, the half disturb output is reduced in amplitude, but this results in a longer cycle time.

FOUR-μSEC COINCIDENT-CURRENT MATERIAL

This material is not of particular value in high-speed memories as coincident-current cycles using this material must be about 20 μsec in duration. It can, however, be used in linear selection circuits to built economical transistor driven memories having cycles in the 8- to 10-μsec region.

In the coincident-current mode the material must be driven with pulses which are at least 6 μsec long at the top. The full current employed is about 220 ma. The disturbed ONE output is about 16 mv and the others are correspondingly small.

In a linear selection mode a read pulse of 2-μsec duration and 350-ma amplitude is employed. The ONE signal is 50 mv and the disturbed ZERO is 4 to 5 mv. Writing is accomplished in 4 μsec with about 220-ma drive. The write signal are small and therefore a large number of cores can be driven.

Requirements of the Transistor in the Driver Circuit

Among the most difficult problems which faces the designer of a high-speed core memory are those associated with the transistor driver circuits. Magnetic cores with short switching times require high drive currents which rise rapidly. Since the load is inductive, this places severe demands upon the forward voltage rating of the driving transistors. All three of these requirements are incompatible and tend to increase the difficulty in selecting a suitable transistor of a readily available type and at a reasonable cost.

Experiments have been performed which show that the turn-on switching time in a transistor increases approximately linearly with the current being switched. Measurements were taken in the region of 100 to 500 ma. The transistor must exhibit extremely fast switching characteristics in this range of currents. Unfortunately, transistors with faster switching speeds generally have lower collector-voltage ratings. This means that fewer cores can be driven by the transistor circuit.

These difficulties have been alleviated to some extent by the use of the 50-mil cores. It seems very unlikely that cores smaller than 50 mils can be handled by present assembly techniques. Higher speed memories of the future may consist of faster transistors and smaller "cores" in the form of thin films or twistors or storage elements of a lower energy level, such as the cryogenic device.

CONSIDERATIONS LEADING TO THE DRIVER CIRCUIT

The functions of the current driving circuit are to supply a well defined constant

Vol.
Pt. 1
195

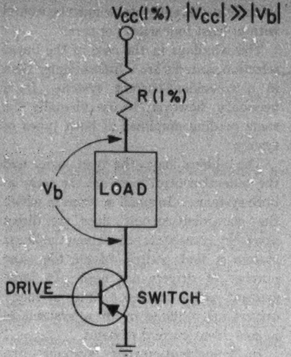


Fig. 1. Transistor switch used with precision resistor

drive current to the load regardless of the back emf (electromotive force) generated within the load, and to turn this current completely on or off with switching times within the requirements of the cores being driven.

One form of current driver makes use of the common-base transistor connection with its nearly ideal constant current output characteristics. Serious limitations exist in this type of circuit.

A considerable voltage must exist across the transistor if a wide variation of voltage drop in the load is to be accommodated. At the high current involved the power dissipation is high enough to require the use of a power transistor. Power transistors are relatively slow in switching times. Another disadvantage is that a large emitter current, which is larger than the collector current, must be switched at the input to the driver.

Another form of driver is a common-emitter stage with feedback to control the current. This also requires the use of a power transistor. With present transistors the feedback circuit is too slow to be effective. Probably the most practical approach is to employ a fast switching transistor of the 100-milliwatt class used as a nearly ideal switch. The transistor must be driven well into saturation in order to minimize the power dissipation.

In addition to the selection of a fast switching transistor, much can be done to enhance the speed of the switching circuit. It has been established[1] that a transistor can be turned on more quickly if overdriven and that the storage and turn-off times can be reduced by increasing the turn-off drive. Overdriving gives rise to an additional advantage of lowered saturation voltage. If the base can be slightly reverse biased when the transistor is to be held at cutoff, a somewhat higher collector voltage may be used. When it is inconvenient to use reverse bias the resistance of the input circuit from base to emitter must be kept as low as possible to avoid the "open base" condition.

The Driver Circuit

The highest switching speeds of currents in the several hundred ma region are obtained with small transistors used as nearly ideal switches. As previously discussed, the transistor must be driven well into saturation to keep the saturation voltage as low as possible and must be overdriven in both directions to minimize the switching times.

The transistor switch is a poor current source. Some other means must be provided to maintain a constant drive current. When the back emf V_b generated by the load is very small in relationship to the collector-voltage rating of the transistor, a precision resistor may be used in series with the transistor switch and the load across a regulated power supply, as shown in Fig. 1. This method is used in the inhibit drivers.

If the back emf is an appreciable fraction of the transistor voltage rating V_{cemax} the current-determining resistor may be returned to a much higher voltage providing a clamping diode is used to limit the voltage across the open switch; see Fig. 2.

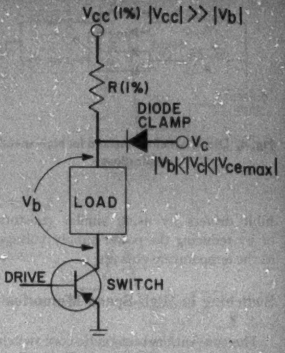


Fig. 2. Transistor switch with diode clamp

This system requires a very high supply voltage V_{cc} and a considerable amount of power must be dissipated in the resistance R. A refinement to this circuit consists of replacing the resistor with a constant current source operated from a much lower voltage.

The constant current source uses a power transistor in the common-base connection to take advantage of the pentode-type output characteristics. The power transistor is conducting at all times and receives constant current through either the clamping diode or the load. It is not required to switch the relatively high current and is therefore not limited by its slow switching speed. The combination of the constant current circuit and the clamping diode is called the current stabilizer. Two types of current stabilizers have been developed. One supplies a constant negative current and the other a constant positive current. They are called the positive current stabilizer and the negative current stabilizer.

Both of these current stabilizers are required for the bidirectional driving of the cores on a single set of windings as illustrated in (Fig. 3). The input to the transistor switch is transformer coupled so that both terminals of the switch may be made floating and so that they may be connected in either direction.

Many transistor switches may be connected to a single current stabilizer if only one of the switches is to be energized at a given time, which is the case in the operation a core memory system.

Temperature Compensation in the Current Drivers

As the temperature increases in a ferrite core memory matrix, the required coercive force of the core decreases for optimum operation. This is particularly true in a coincident-current core memory. The drive currents must be reduced as the temperature increases if a wide operating temperature range is desired. Core memories have been operated over a temperature range in excess of 80 F by using temperature compensation.

The current stabilizers have current-defining resistors with the proper positive temperature coefficient. The resistors are wound with Belco wire. Temperature compensation can be achieved in the in-

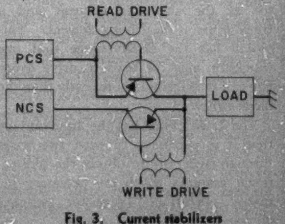


Fig. 3. Current stabilizers

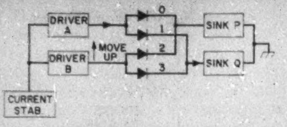


Fig. 4. Diode switch connected for high-speed applications

hibit drivers by using similar resistors or by reducing the power supply voltage as the temperature goes up.

Switching in High-Speed Memories

The conventional magnetic-core switch has considerable combinatorial advantages, but these are offset by the low overall efficiency of the switch. At high switching rates the inductance of the switch introduces voltages which are much too high to be withstood by transistors, and no form of conventional magnetic-core switch has been found which is suitable for high-speed transistor-driven applications. It is possible to construct a switch using linear transformers and diodes. This is sufficiently efficient for use in high-speed applications. However, the voltage and current levels required by the matrix stack is usually the desired operating levels of the transistor and therefore the linear transformer serves no useful purpose.

A simple form of diode switch can be employed. The manner in which this switch is connected into this system is illustrated in Fig. 4.

For each polarity, the system employs n drivers, denoted by A, B, etc., and m current sinks, denoted by P, Q, etc. Such a system can drive $m.n$ matrix lines and requires the use of $m.n$ diodes. In the illustration, current is routed from the current stabilizer through driver A, then through driver Q to select line 1. The diode employed must be capable of withstanding the full voltage and current appearing at the matrix lines.

The simplest form of switch is that illustrated in Fig. 3, which uses a transistor as the switching element and a transformer as the driving element. This switch has the disadvantage that one transistor must be used to drive every line and is, therefore, more costly than the diode switch.

The diode switch has the disadvantage that it is necessary to wind two wires through each core to obtain the bidirectional currents which are necessary for matrix operation. Transistor switches do not have this disadvantage.

ADVANTAGES OF LINEAR SELECTION SYSTEMS

Linear selection operation of the memory allows the use of very short pulses; consequently, linear selection systems are inherently rapid in operation. The propagation time through the system is negligible, as only as many cores as there are bits in the word are driven at any one time. The signal-noise ratio from linear selection systems is not usually as good as from an ideal coincident-current system but, on the other hand, the signals are clear and stable and the reading operation is usually simple to perform.

The power required to drive a linear selection system is low because the number of cores that are driven is small

DISADVANTAGES OF THE LINEAR SELECTION SYSTEMS

The principal disadvantages of the linear selection system lies in the fact that it is necessary to drive as many lines as there are words in the memory. This makes the use of a diode switch almost completely essential. In medium and large size memories this diode switch introduces severe problems as leakage in diodes and recovery time make it impossible to connect very many diodes to each driver. Experiments have shown that over a reasonable temperature range, and at high speeds of operation, it is not advisable to connect more than 32 diodes to any one driver.

In a 4,096-word system it is necessary to employ 8,192 diodes, half of which must be capable of carrying up to 500 ma peak current. Theoretically, 192 drivers could be used to drive the diode switch. In practice, however, it is necessary to use 512 drivers to overcome diode leakage and recovery problems. This places a limit on linear selection systems in that they tend to become uneconomical at sizes greater than 1,024 words.

One further disadvantage of the diode switch is that it is necessay to time the drivers rather carefully because the operation of the driver and the current sink must be relatively well timed together. This close accuracy of timing is difficult to achieve in large systems.

COMPARISON BETWEEN LINEAR SELECTION AND COINCIDENT-CURRENT SYSTEMS

It is necessary to use a crudely compensated sense winding in linear selection systems. This is required to minimize noise during reading and also to reduce voltage excursions in the reading amplifier during the writing operation. For this reason both linear selection and coincident-current systems must be wound with at least four wires per core.

The windings in the case of the linear selection systems are a little simpler than in a coincident-current system. It is necessary, however, to use virtually the same reading amplifier in both types of system.

The address logic, the read logic, and the general control logic are the same in both systems. In small systems in which the coincident-current lines are direct driven by transistors or in which the linear system is diode switch driven, the same number of drivers are used. In large systems, twice or even 4 times as many drivers are required in linear systems as in coincident-current systems.

In coincident-current systems one driver type can be used to perform all drive functions; this leads to economy of manufacture and design. In a 4-μsec-cycle system this driver must deliver about 250 ma at 10 volts. In linear systems it is necessary to employ at least four types of drivers. In a 3-μsec system these must be: (a) 400-ma driver, (b) 400-ma sink, (c) 100-ma driver, (d) 100 ma sink. In some linear systems it is necessary to add a fifth type of driver which is used in the digit drive circuit.

In general, linear selection systems must be employed for cycle times below 4 μsec. Above 4 μsec there is a choice between coincident-current and linear systems. Up to 1,024 words of storage, coincident-current systems are generally more efficient for short or medium length words. When word lengths exceed 40 to 80 bits, depending upon speed, it is usually advantageous to employ a linear selection system. Above 1,024 words, even if the word length is very long, coincident-current systems are, in most cases, more efficient than linear selection systems.

Possible Transistor-Driven Memories

The general techniques described in this paper can be used to design highly efficient transistor-driven systems which are either very small and very fast or relatively large and relatively slow. Development within the scope of this paper has been aimed at producing 3 series of modular memory units.

First, a memory with a 3-μsec total cycle time which stores either 100 or 128 words of up to 80 bits each. Second, a series of memories having capacity between 1,000 and 8,192 words of length up to 80 bits each. These memories have cycle times between 6 and 8

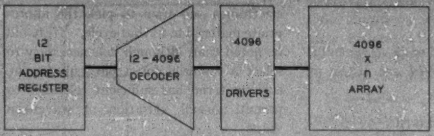


Fig. 5. Diagram for 4,096-word memory, n = number of bits in word

μsec, depending upon the total capacity. A third series of memory can be constructed using these techniques. This has a capacity of 1,024 words or less, of up to 40 bit words with a 4-μsec cycle. The first memory employs linear selection techniques, while the second and third employ coincident-current techniques throughout.

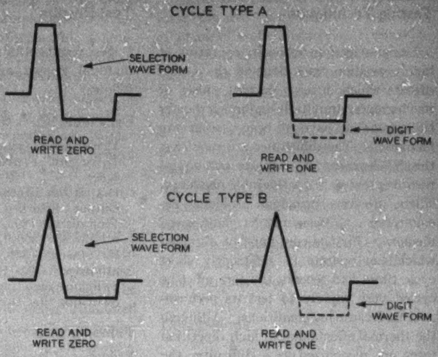


Fig. 6 (right). Typical cycles

Addendum: The Behavior of Ferrite Cores at Very High Operating Rates

The operation of low-speed ferrite-core memories is relatively insensitive to changes of pulse shape. It has been determined experimentally that cores operated at high repetition rates are sensitive to pulse shape and the optimum material for a given wave shape must be chosen with some care. Cores which operate successfully with short pulses of poor wave shape will not operate successfully with pulses of the same length, but of rectangular shape. The converse is also true. The paper describes experiments that have been performed and postulates a reason for this phenomenon.

Introduction

It has been shown in many recent papers that the linear selection, or word selection, or end-fire technique, is most suitable for operating ferrite cores at high repetititon rates. Basically, this technique rejects the classical coincident-current mode and instead employs one selection wire for each word in the memory. It is necessary to employ some driver and decoding arrangement between the address register and the core array which will make it possible to select an individual word in the memory. This is illustrated in diagrammatic form for a 4,096-word memory in Fig. 5.

In practice, the drivers are introduced into the decoding matrix so that the number of drivers employed are usually no more than are used in a conventional coincident-current memory. This technique has the advantage that propagation delays associated with coincident-current memories are eliminated.

It places a much greater burden on the electronic decoding, but the use of present transistors and diodes makes it possible to perform the decoding operation rapidly. This technique also has the advantage that a new type of cycle may be used. This cycle is a characteristic feature of all linear selection systems. Two typical cycles are illustrated in Fig. 6.

Cycle-type *A* is characterized by very rapid rates of rise and fall of the current during the positive pulse. This pulse is the read pulse of the cycle. The negative pulse, or write pulse, may or may not have rapid rise and fall times but is always much smaller and longer than the positive read pulse. A ONE is inserted into a particular digit by energizing a digit drive winding. The current in this waveform is usually about two-thirds of the current available in the write selection waveform. It is clear that the normal 2-to-1 relationship of a coincident-current memory is not employed in a linear selection waveform.

In cycle-type *B* the read waveform is triangular and consists of the lowest rate of rise and fall consistent with obtaining a pulse no longer than the rectangular waveform employed in cycle *A*. The peak current in this triangular waveform is usually somewhat greater than the peak current in the equivalent rectangular waveform. The rules governing the write waveforms are the same as those in cycle-type *A*.

In practice, a large variety of waveforms can be used between these two extremes and frequently a half sine wave is used for read, and a long, low amplitude half sine wave for write. In high-speed memories the currents employed are all always quite large and it is very important to use a cycle in which the rise times are as long as possible so that relatively conventional transistors and diodes can be employed in the decoding matrices.

The output signals from any linear selection system consist of a large amplitude, short duration pulse for a read ONE, and a low amplitude, short duration pulse for a read ZERO. The write ZERO pulse is usually short and of negligible amplitude while the write ONE pulse is usually relatively very long and frequently of no greater amplitude than the read ZERO pulse. This simplifies the problem of sensing in linear selection systems as only a read ONE pulse is of large amplitude.

The principal advantages of a linear selection system are threefold. First, the pulses employed are very short; therefore, very fast cycle rates are possible. Second, only a small portion of the total flux available in the core is employed and the heat dissipation within the core is reduced considerably. This allows the core to be operated at high rates without encountering serious overheating effects. Third, the cycle employed permits the use of cores which do not have the same characteristics as those employed in coincident-current systems. It is not immediately apparent that this is an advantage. All cores employed in coincident-current systems tend to fall on or near a curve which has the characteristics that increasing currents must be supplied to the core as shorter switching times are desired. It has been found that cores which can operate successfully on a linear selection cycle can be moved off this curve so that switching times may be reduced as much as 4 times for a given current. These cores will not operate successfully in a coincident-current memory.

Testing Techniques

Linear selection memories operating at high repetition rates demand the use of drivers which develop current pulses of much greater amplitude and much shorter rise times than have yet been used in core memories. The difficulties are such that the development of adequate test equipment represents a considerable challenge. There are two principal problems encountered in designing a high-speed memory. First, a core must be obtained which can operate at sufficiently short cycle times and second, the core must be cycled very rapidly to test its performance under heating conditions. Although the thermal effects due to high repetition rates are recognized as a difficulty, the major preoccupation is that of obtaining a core which can be read and written in a very short time. Mercury relays combined with transmission lines represent a simple approach to the problem of generating short cycles of the required type. The cycles generated, however, are almost always of type *A* in Fig. 6, as effective, instantaneous rise times are available from mercury relays. There is a trend to perform all initial tests with near ideal pulses. In very small memories or in certain special types of memories, such as sequential memories, it is possible to develop drivers which will generate these near ideal pulses economically. In a random access memory it is essential that the least ideal pulse should be used, as this will have a large effect upon the economy of the decoding system.

It is both convenient and prudent to design a tester for high-speed checking which has the characteristics of cycle-type *B* in Fig. 6, as this cycle is the one that is most likely to be encountered in practice. If high-speed cores are tested on two dissimilar testers, it is found that it is not possible to correlate results obtained on both testers. Only when a short-cycle-time low-repetition-rate tester is designed to have the same characteristics as the high-rate tester is it possible to correlate performance of cores.

It has been found that not only do cores change their characteristics as heating occurs, but that, in addition, the cores are relatively sensitive to the type of cycle that is employed.

From the preliminary results that have been obtained it is obvious that testers must be used in which the complete waveform to be employed in the final memory can be simulated in testers of all types. Only if this precaution is taken can the results of testers be trusted.

Test Results

Test results obtained on developmental material known as *6914* are listed as follows:

OPERATING CHARACTERISTICS

Drive current (read) ma	500
Drive current (write) ma	280
Knee	165
Output (undisturbed ONE) mv	100
Output (disturbed ZERO) mv	16
First disturb (300/180) mv	2
*n*th disturb (300/150) mv	2
First disturb (300/180) mv	3
*n*th disturb (300/180) mv	3
Switching time (read) μsec	0.3
Switching time (write) μsec	0.4

THERMAL EFFECTS

Reduction of knee ma	
20-μsec cycle	2
10-μsec cycle	4
5-μsec cycle	7
2.5-μsec cycle	9
1.5-μsec cycle	13

EFFECT OF VARIATION OF WRITE HALF CURRENT USING 1.5-μSEC CYCLE

dV2 mv	
127 ma	20
147 ma	25
163 ma	30
167 ma	35

These results were taken using a cycle such as type *B* and it is seen that under these conditions this core is an effective storage device.

If the *6194* is tested using a cycle such as type *A*, it is found that the read ZERO signal is almost as large as the read ONE but of very short duration. In using various cycles between these two extremes it is found that in the case of this particular material the read ZERO signal is a direct function of the rise time of the read pulse. The material behaves extremely well if the rise time is sufficiently long, but very poor results are obtained on a short rise time.

Conversely a classical rectangular loop material can be tested. In this case it is found that the performance under cycle-type *A* conditions is very good. A large short read ONE signal is obtained and a small short read ZERO signal is obtained. However, when this material is tested under cycle-type *B* conditions, the read ZERO signal is reduced somewhat but the read ONE signal is reduced materially.

A large number of materials of various types have been checked and it has been found that they cover a whole gamut of characteristics between the two extremes that have been quoted. From this it is concluded that in designing a core memory it is necessary first to select the waveform which can be achieved by the electronics and then to pick the appropriate material to fit that cycle. It should be noted that, although it is necessary to select a material for a particular cycle, each material has sufficient tolerance to permit normal variation of waveform shape.

In performing these tests it was noted that the *6914* material was least sensitive to heating under rapid cycling conditions and, in general, the cores became more sensitive to heating as the rectangularity of the core increased.

At present *6914* appears to be the best material for memories of high repetition rates. It tolerates the slowest possible rise times consistent with obtaining short cycles, and also exhibits the best thermal properties.

Explanation of Phenomenon

The hypothesis developed in this section is almost certainly not correct. It does, however, provide a model by which most of the effects observed can be explained.

It is known that nucleation is an important factor in the operation of a ferrite core. It has been shown, both experimentally and theoretically, that square loop materials nucleate from relatively dispersed points and that domain wall movement then proceeds from these points. The number of points at which nucleation occurs is a function of the magneto-motive force. The higher the value of *H* the more points of nucleation. It is considered, therefore, that a rectangular loop ferrite core tends to switch, starting at the inside diameter where *H* is greatest, and then proceeds outward to the outside diameter where *H* is least.

In a linear selection operation the read current employed is much larger than the normal switching current and there is probably very active nucleation around the inside diameter; the domains then travel outward in the form of a wave toward the outside diameter. If the pulse employed is very short this action becomes interrupted so that only the inner annulus of the core is switched from

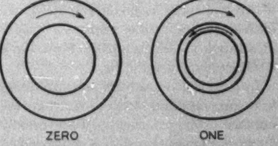


Fig. 7. Inner annulus switched from one state to another

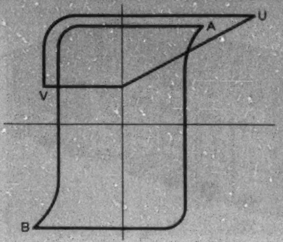


Fig. 8. Typical d-c loops using same core

one state to another. This is illustrated in Fig. 7.

During a write action this annulus is returned to its normal state so that the whole core is magnetized in one direction. This action proceeds slowly as the currents are much lower, but exists for a longer time. This action tends to maintain the bulk of the core in one direction of magnetization at all times.

Under this mode of operation only a small portion of the core is used and, therefore, the operation is analogous to employing a core that is extremely thin walled and of mean diameter, equal to the inside diameter of the core.

It is believed that an operation of approximately this type occurs when cycle-type *A* is employed. This is substantiated by the facts that the degree of magnetization of the core and the electrical performance of the core is a function of all aspects of the pulse employed. Relatively small variations of rise time, pulse duration, and amplitude, produce large variations in performance. This is to be expected if the operation employs relatively unstable flux conditions.

It is believed that another effect is present. It is possible to switch a core partially under d-c conditions. Under these conditions it is difficult to believe that only an annulus is reversed. Fig. 8 illustrates two typical d-c loops taken using the same core.

Loop *AB* is a d-c hysteresis loop of a conventional rectangular material tested under simulated coincident-current conditions. Loop *UV* is a minor loop of the same material tested under simulated linear selection conditions. Loops similar to *UV* but much smaller in all respects, are obtained under d-c conditions using *6914* material. It will be seen that the read half of the loop is that associated with point *U* and that in this part of the loop no form of coincident-current operation is possible; but if the core is pulsed with a short pulse, very rapid operation will result as permeability is very low. However, on the other half of the loop associated with point *V*, a conventional square loop form is available, and in this area the coincident-current-type operation associated with writing linear selection systems can be operated quite successfully. Material *6914* was developed using the hypothesis that a bulk effect took place in the material and that the high-speed performance of the material could be related to its d-c performance.

There are, therefore two possibilities. First, the effect may be due to switching of an annulus of the material or, secondly, the effect could be due to switching of the bulk of the material, but on a minor loop. In practice, of course, the effect is likely to be a combination of both of these. It has been found that operation of *6914* using a *B*-type cycle, is more uniform and stable than operation of rectangular loop materials using *A*-type cycle; and this would seem to confirm the theory that the effect in the *6914* material is more of a bulk effect than an annulus effect, while in rectangular loop materials the effect is more annulus than bulk.

Conclusions

The effects within the body of a ferrite core are a function of the material and of the nature of the waveforms imposed upon the material. Optimum results can be obtained only by selecting the material which most naturally fits the cycle employed. In addition, it is necessary to to employ accurate simulating techniques when testing materials as it is possible to obtain completely erroneous results by employing a waveform which differs from that of the memory under design.

Reference

1. LARGE-SIGNAL TRANSIENT RESPONSE OF JUNCTION TRANSISTORS, J. L. Moll. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 42, 1954, pp. 1773–84.

A New Transistorized Negative-Impedance Telephone Repeater

R. P. DIMMER
NONMEMBER AIEE

E. L. ROBACK
NONMEMBER AIEE

WITHIN the last decade, the use of the negative impedance type of repeater in the telephone plant has proved to be a very satisfactory device to improve transmission on circuits up to about 25 miles in length. This type of repeater is both simple and low in initial cost, and when properly installed, its operating and maintenenace cost is practically negligible.

Of the two types of negative impedance repeaters available, the series type was the first to be used. This repeater adds a negative impedance in series with the transmission medium which in effect reduces the loss of the circuit. Although this type usually provides sufficient gain for most applications, it does produce some undesirable transmission characteristics such as echoes, particularly noticeable when used on multiswitched circuits. In order to reduce these effects, a shunt negative-impedance repeater was designed later to be used in combination with the series. The use of the shunt repeater also provides some additional gain. In most cases today, the use of the combination series-shunt repeater will be required to meet present-day design objectives. With both repeaters working as a team, the transmission engineer now has an excellent tool with which he can reduce the loss of most circuits satisfactorily at a cost within reason.

The extent of the application of the combination repeaters becomes largely a matter of economics. One very important economical factor which governs its use is the difficulty in line-up sometimes encountered. When facilities are not in good order, considerable restrapping may

---

Paper 59-794, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 20, 1959; made available for printing May 4, 1959.

R. P. DIMMER and E. L. ROBACK are with General Telephone Laboratories, Northlake, Ill.

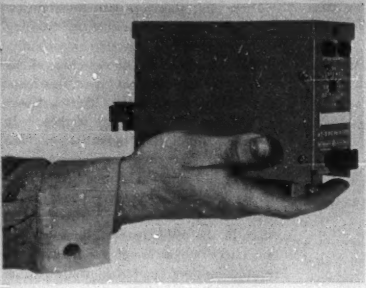


Fig. 1 (left). Type AT-5 repeater

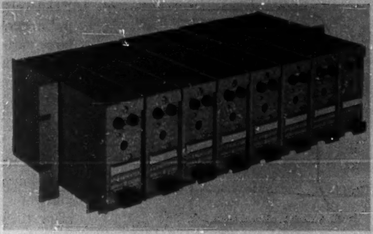


Fig. 2 (right). AT-5 repeaters mounted on shelf

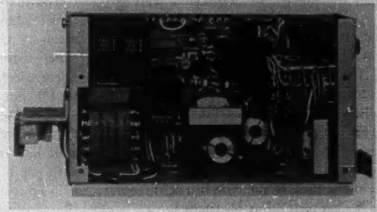


Fig. 3 (right). Interior of AT-5 repeater

be required to obtain stable repeater operation. This restrapping takes time, and requires personnel with some knowledge of strapping technique.

Present negative impedance repeaters have network components made up of many graded values of resistance, capacitance, and inductance. These are trapped in the field in a definite configuration to produce a desired gain. Many of these network components are not always used at every instance; consequently a user is penalized in cost for the parts not required. Precision components are also required to obtain reproducible gains from a number of repeaters on like facilities.

In view of these problems, it would be very desirable if a negative impedance repeater were available in which the gain could be changed by a single control, necessitating little or no strapping and which could handle most of the facilities in use today. Such a gain control technique should require the use of but a single fixed network. This is especially important today where such repeaters are placed into service by personnel unskilled in the techniques of transmission. With these objectives in mind, the General Telephone Laboratory designed a combination series-shunt negative impedance repeater, known as the Type *AT-5*, which utilizes an adjustable control on the front panel to vary the gain of the unit. This new repeater is shown in Figs. 1, 2, and 3.

Theory of Operation

Negative impedance repeaters of either type are generally composed of three parts; namely:

1. Coupling section.
2. Amplifier or converter section.
3. Network section.

The coupling section comprises either a transformer or capacitors to isolate the amplifier section from the line except for its intended voice frequency gain effects. The amplifier section comprises a fixed amplifier circuit using either vacuum tubes or transistors in either a single ended or push-pull configuration. The network section is composed of various resistors, capacitors, and inductors which can be strapped in a definite configuration to produce a required gain or equalization on the transmission medium. A block diagram of a combination repeater connected into a line between two offices is shown in Fig. 4. Note that the series repeater is placed in series with the transmission circuit, and the shunt repeater section is connected across the midpoint of the series unit in order to produce equal effects to both sides of the repeater.

A negative impedance repeater applies a negative impedance or admittance to a transmission medium as a result of a conversion of a positive impedance within the amplifier section. This positive impedance is the network; the amplifier section is the converter. Normally the

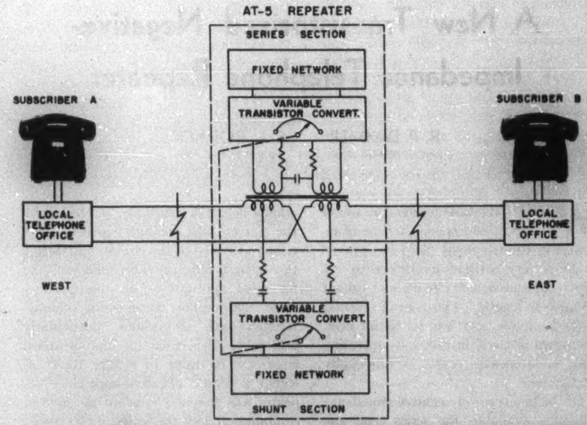


Fig. 4. Block diagram of combination repeater AT-5

Vol. 7
Pt. 1
1959

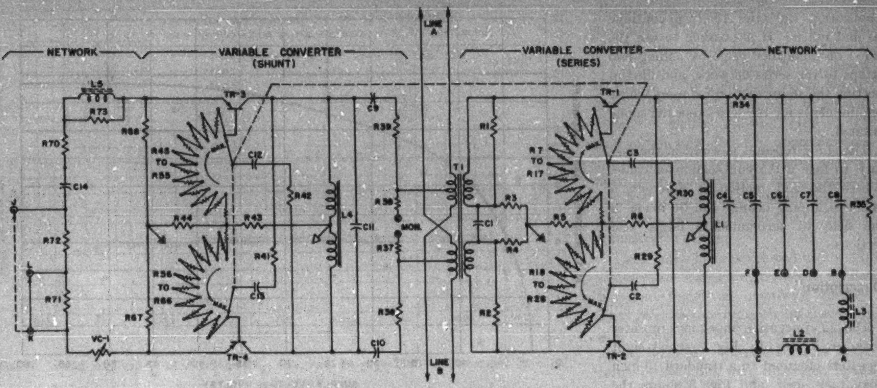


Fig. 5. Circuit of AT-5 repeater

variation of the network impedance controls the amount of negative impedance or admittance introduced on the telephone circuit. In order to use network components of reasonable size, the converter coupling section changes the effect of the impedance of the network by a fixed high-to-low ratio. In the Automatic Electric Type *AT-2* transistorized repeater, the ratio of the converter is approximately 10–1. For example; if a resistance of 10,000 ohms is connected to the network terminals, a negative impedance of approximately 1,000 ohms is reflected in series with the transmission line. This conversion is made by the amplifier section of the *AT-2*. It is also possible to use a converter with a 1-to-1 ratio, and change the ratio to 10-to-1 in the coupling device. The best method is dictated by impedance considerations. In either case the amplifier section plus the coupling device presents a constant conversion ratio factor to the network. The network is the variable element, and by changing the value of the components in it, one can obtain a variable negative impedance or gain on the transmission medium. To get various degrees of gain the strapping of many graded components is necessary. One way to eliminate strapping would be to design a repeater with all network elements variable; such a unit however, would probably be many times greater in size and undoubtedly excessive in cost. Another approach would be to use a fixed network and a variable converter. The *AT-5* repeater utilizes this principle as shown in Fig. 5.

To prove the feasibility of this method, a simple technique of varying the conversion ratio of the amplifier section of the repeater had to be found. In the *AT-2* or *AT-3* the conversion ratio of 10-to-1 is obtained by the ratio of feedback voltage on both sides of either repeater. By making these elements variable, a variable converter is obtained. This technique is shown in the circuit of the *AT-5* (Fig. 5). The gain of the various steps using this technique is shown in Fig. 6, and is very similar to that produce with a separate *AT-2* and *AT-3* combination repeater. This family of curves shows the gain steps up to the maximum of 9.8 decibels (db); however, gains above 7.0 db are not recommended in terminal applications.

To make such a repeater practical certain design assumptions have to be made; namely:

1. All cable will be loaded.
2. All loaded cable will be of the *H-88* type (88 millihenrys every 6,000 feet).
3. The end section facing the repeater will be 0.5 section.
4. It should require only one type of repeater for the most widely used gages of cable without changes or strapping.

If these assumptions are correct, then the *AT-5* will provide a voice repeater where the gain of a circuit can be changed by the simple adjustment of a front panel switch.

In most cases of transmission practice, cable will be loaded where possible before adding a repeater. In those cases where nonloaded cable is present, particularly on foreign service lines, a shunt negative impedance section cannot be used to advantage; consequently a combination repeater is seldom considered. In these

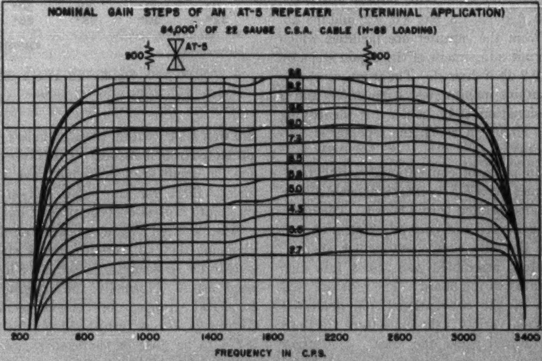


Fig. 6. Transmission characteristics of AT-5 at all gain steps

applications the series *AT-2* will continue to find usage. It is possible to use the *AT-5* on nonloaded cable by adding two straps to the series network, and disconnecting the shunt section for users whose requirements for the series repeater alone is small.

The *AT-5* repeater is wired in the factory to face a 0.5 end section of cable. Where the end section is different, a minor strapping change can be made in the field to accommodate larger end sections.

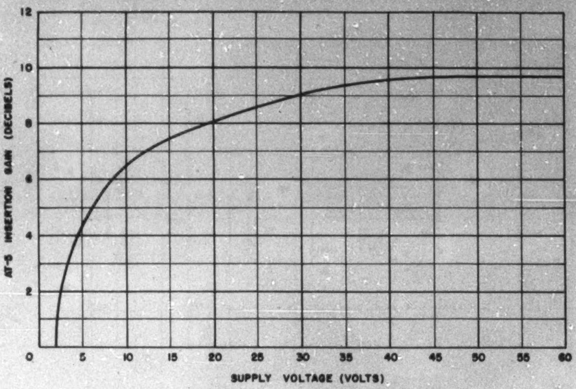


Fig. 7. Effect of supply voltage change on AT-5

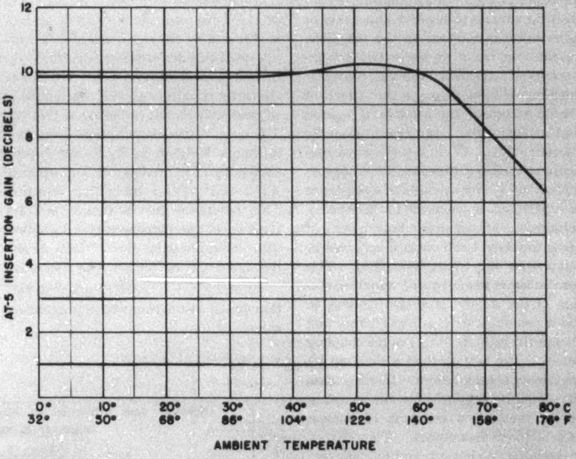


Fig. 8. Effect of ambient temperature change on AT-5

Description

The *AT-5* negative impedance repeater is shown in Fig. 1. Fig. 2 shows eight repeaters mounted on a standard 19-inch relay rack shelf, and Fig. 3 shows the interior of the unit. The unit is 2 inches wide and 9$^1/_2$ inches deep, and occupies a shelf height of 5$^1/_4$ inches. Up to eight repeaters can be mounted on a single shelf for 19-inch relay rack, or ten repeaters can be mounted on a shelf for the 23-inch rack. Each repeater is designed to slide on a track to engage an 8-point shelf jack of a type long familiar with Automatic Electric telephone equipment. These jacks are adjusted at the factory to switch both sides of the line straight through when a repeater is removed. This very desirable factor permits rapid replacement of a defective repeater if trouble should occur, without interrupting an existing conversation. The repeater jack and shelf arrangement are the same as used for the Automatic Electric Company *AT-2* and *AT-3* repeaters. Where a circuit is wired for an *AT-2*, the *AT-5* can be used directly without any changes.

The *AT-5* has an adjustable switch on the front panel which permits adjustment of gain in the following steps: 2.7, 3.5, 4.3, 5.0, 5.8, 6.5, 7.3, 8.0, 8.6, 9.2, and 9.8 db. Within the repeater are facilities which permit adjustment of these gain steps by ±0.2 db if desired. The front panel also contains monitor test connections and a designation strip holder for identifying the circuit in which the repeater is connected.

In place of conventional wiring techniques the *AT-5* repeater is assembled on a 2-sided printed wiring board. The input transformer and gain adjustment switch are the only components which are separately assembled and wired into this printed wiring board. The resistors used are of the conventional 1% deposited carbon variety; the capacitors are both metallized paper and mica. The larger capacitors are contained in tubular cylinders and mounted vertically. They are well sealed within these cylinders and protected from moisture. The transistors for the shunt section do not have heat sinks; those for the series section incorporate a press-on type of heat sink.

The printed wiring board permits changing the end section network strapping in the series unit from 0.5 to 0.8 if the need arises. Normally the repeater is furnished strapped for 0.5 end section. Terminals are also furnished to enable the *AT-5* to be used on nonloaded cable if required by adding two straps, and removing two straps within the unit.

Performance

This repeater requires no power source other than the conventional 48-volt telephone exchange battery. The unit consumes approximately 27 milliamperes at 48 volts (approximately 1.3 watts); consequently, heat radiation is not a factor as is found in vacuum-tube repeaters. The unit is not critical with changes in supply voltage as shown in Fig. 7. Between 38 and 60 volts, the gain of the unit will change only 0.2 db from its nominal value at the highest 9.8-db setting.

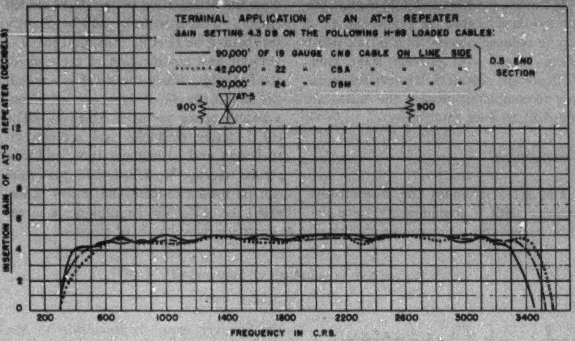


Fig. 9. Field application of AT-5 on 19-, 22-, and 24-gage H88 loaded cable (terminal application)

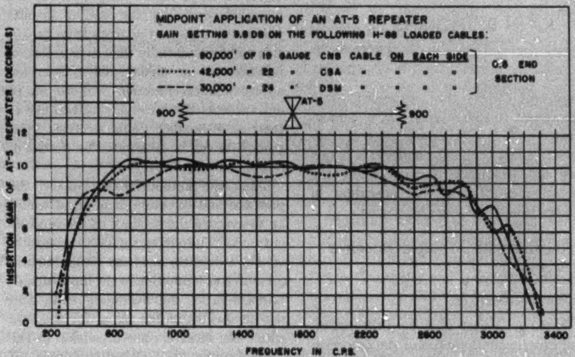


Fig. 10. Field application of AT-5 on 19-, 22-, and 24-gage H88 loaded cable (midpoint application)

The *AT-5* will drop to its 1/2 gain point with a supply voltage of 6.0 volts.

In all applications this type of repeater provides its maximum gain when installed at the middle-third of a line. The *AT-5* will provide gain equal to about 60% of the loss of the line when installed as a terminal repeater with 7.0 db as maximum. In midpoint application, gain equal to as much as 75 to 80% of the loss of the line can be utilized with 10.0 db as maximum. In order to take advantage of this midpoint characteristic, it is possible to install these units along a line (possibly mounted on poles), and supply them with voltage from the office exchange battery through a spare pair in the cable. Transistorized repeaters have a very significant merit for such applications, in that changes in supply voltage cause only small changes in the negative impedance that is presented at the input terminals. Reduction in supply voltage effects only the output (overload) capabilities of the repeater. This effect generally causes no difficulty because as the repeater is installed further from an office, its input signal is also reduced; consequently over-loading rarely occurs. The *AT-5* is capable of handling outputs of over 10 volume units before compression or distortion occurs.

Since the transistors used in the *AT-5* repeaters are of the germanium variety which are sensitive to temperature changes, the effect of high ambient temperatures must be considered. This particularly is a factor if the repeaters are to be pole mounted, or used in extreme hot climates where air conditioning is not available. The gain in the present design remains constant to 40 degrees centigrade (C) or 104 degrees Fahrenheit (F), shown in Fig. 8. The gain rises about 5% at 50 C (122 F) and drops back to normal at 65 C (149 F). At 85 C (185 F) the gain of the unit drops to approximately 1/2 of its normal value. In most applications these changes are of no significance, but in extremely hot ambient temperatures, it is well to consider this fact in the transmission design.

These units have developed no trouble under extreme high line transient, similar to those found on open-wire circuits. They have been tested with line transients in the range of 1,000 volts, and showed no sign of damage. Ringing voltages in the range of 200 volts were connected to the repeater under the most severe trouble conditions and produced no apparent change in performance.

Application

Fig. 9 shows the insertion gain of an *AT-5* when used as a terminal repeater on three gages of *H88* loaded cable, 19 *CNB*, 22 *CSA*, and 24 *DNB*. These curves are for a gain setting of 4.3 db, the maximum recommended for the facilities shown. The repeater provides approximately the same gain for each of the 3 different sizes of cable, except for a faster drop in gain at the upper frequency end on the 19 gage size. The response of Fig. 9 appears quite smooth with a resistance termination, however, when connected to phones in a normal use condition, the usual hills and valleys will appear.

Fig. 10 shows the same repeater in a midpoint application on longer circuit. In this application the *AT-5* repeater can provide gain equal to about 75% of the loss of the line up to 10 db. Here the impedance irregularities show up to a greater degree.

Installation

The installation techniques used for a *AT-5* repeater are extremely simple. Assuming the facilities are in good order, standard transmission design practices would be first considered in order to determine the maximum gain that can be inserted into the particular facility by the use of a negative impedance type of repeater. When this is determined, the adjustable control on the unit is set, and the repeater is ready for use. This assumes that the end section of the cable facing the repeater is 0.5. If this is not the case, a strap can be changed inside the unit to permit operation between 0.5 and 0.8 end section. In most cases a short- and open-circuit stability check would be made at the closest terminal. If the facility is satisfactory, the repeater would be stable for the gain set. If it is desired to refine the nominal gain of the repeater by a factor of ±0.2 db from the gain control values, provision is provided within the unit.

Because of the ease of setting gains on the *AT-5*, the singing point margin can be quickly determined.

If trouble conditions should change the return loss of a cable, the gain of the *AT-5* can be reduced quickly to a usable point where it will still provide some gain. With present repeaters, it is usually necessary to completely remove the unit, thus affording no gain whatsoever.

Conclusions

The *AT-5* provides the transmission engineer with a simplified negative impedance repeater which can be used to reduce the attenuation of any *H-88* loaded circuit. This new design can be directly substituted on circuits utilizing the *AT-2* without any modification of the rack or shelf wiring. Since the *AT-5* is one-half the size of the *AT-23* combination a considerable space saving is attained. The *AT-5* design simplifies and reduces installation and adjustment time, and as a result, average unskilled field personnel can be used to install the units. Likewise, the *AT-5* design provides one repeater for many types of facilities.

An Experimental Study of Detection in Nonstationary Noise

THOMAS R. WILLIAMS
NONMEMBER AIEE

JOHN B. THOMAS
MEMBER AIEE

THE PERFORMANCE of any system for detecting a noise-like signal in additive background noise is critically dependent on the statistics of the signal and the noise. In analyzing the performance of a given detector it is necessary to make assumptions regarding these statistics. The situations in which both the noise and the signal are Gaussian and stationary (or very nearly so) are quite numerous, and the assumption that such is the case has led to useful results. There are, however, situations where this assumption is not justified: fading signals and fluctuations in noise intensity are commonly encountered in practice.[1–7] In such cases, for short observation intervals the signal and the noise may often be regarded as Gaussian and approximately stationary and detectors can then be designed and analyzed on this assumption. However, if the observation interval is lengthened sufficiently in an attempt to increase the sensitivity of the detector, the inputs to the detector will eventually be nonstationary and the performance of the detector will differ from that expected for stationary inputs.

A theoretical analysis of the effects of nonstationary noise on the performance of a detection system consisting of an array of receivers, each of which has an output containing a stationary signal component and a nonstationary noise component, has been made recently.[8] In the present paper experimental measurements of the performance of such a system are given and it is shown that the performance is consistent with that expected from theoretical considerations. It is demonstrated that nonstationary noise can adversely affect the performance of a detector, and techniques for improving the performance of detectors operating in nonstationary noise are suggested.

Detection with Multiple Receiver Arrays

STATIONARY INPUTS

In the stationary case, both theoretical and experimental studies of the detection of signals by multiple receiver arrays have been made.[9–11] A review of the results which are pertinent to the present work follows.

Consider two nondirectional receivers (or directional receivers which have similar directivity patterns and which are aligned) which are situated in an isotropic, Gaussian, and stationary noise field; if the separation distance is sufficiently great, there will be negligible cross-correlation between the noise outputs of the two receivers.[10] To simplify the discussion, it will be assumed that the two receivers are constrained to lie at opposite ends of the diameter of a circle in the horizontal plane, and that signal sources can only lie in this plane. Assume that the signal is Gaussian and stationary and that its spectral density has the same shape as that of the noise so that predetection filtering cannot be employed to advantage. The average value of the product of the receiver outputs will be a maximum when the receiver array is broadside to the signal source and approximately zero when the receivers are in line with the signal source, again assuming the receiver separation is sufficiently great. Furthermore, for large separations, a plot of the averaged product of the receiver outputs versus the angle of the array from broadside will be essentially the autocorrelation function of the signal. Of course, the autocorrelation function will be swept out twice as the array is rotated once in the case of nondirectional receivers; this ambiguity can be eliminated by the use of directional receivers.

Thus, one method for determining the presence and direction of a signal would be to average the product of the receiver outputs. The output signal-to-noise ratio of such a detector has been computed and measured as a function of the input signal-to-noise ratio, the averaging time, and the bandwidth of the signal for the case where the receivers are broadside to the signal source.[9–11] Other methods of processing the outputs of the two receivers have been studied as well as the use of more than two receivers.[9–11]

For stationary inputs the ratio of the average signal power to the average noise power at the output of a detector is commonly used as an index of the performance of the detector, and the detection threshold occurs when this ratio is about unity. The situation is much more complex with nonstationary inputs, however, and mere measurement and tabulation of the output

Vol.
Pt.
195

Paper 59-809, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing April 23, 1959.

THOMAS R. WILLIAMS and JOHN B. THOMAS are with Princeton University, Princeton, N. J.

The experimental results reported here were obtained in the course of research sponsored by the Office of Naval Research under contract *N6onr-27005*.

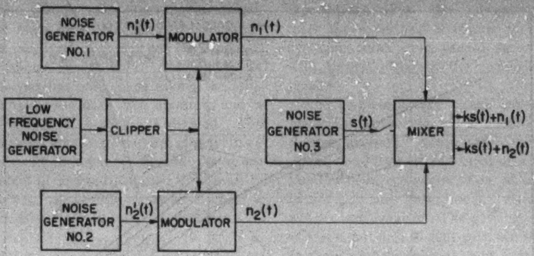


Fig. 1. System used for simulation of receiver outputs

signal-to-noise ratio as a function of the input signal-to-noise ratio does not adequately characterize the performance of a given detector. For this reason the threshold levels reported here were determined by a series of detection trials.

SIMULATED DETECTION PROBLEM

The following problem is considered here: A noise-like signal which is Gaussian and stationary is to be detected using a 2-element receiver array whose outputs are multiplied and averaged. The array is broadside to the signal source. The noise field is Gaussian and isotropic, but the noise at all points in the field is simultaneously modulated by a low-frequency random process. The receivers are sufficiently far apart that the noise in their outputs is uncorrelated, but the modulating envelopes of the two noises are identical. The signal spectral density and the noise spectral density prior to modulation have the same shape.

A block diagram of the system used in simulating the receiver outputs is shown in Fig. 1. Three identical noise generators with band-pass spectra were employed. One served as the signal; the others were amplitude-modulated by a strongly clipped low-frequency noise which was Gaussian and stationary prior to clipping. The signal and the modulated noises were added in suitable proportions in a mixer prior to application to the detector.

For the performance tests the two simulated receiver outputs were fed to a multiplier, the output of which was averaged with an *R-C* averager and displayed on an oscilloscope. The noises at the receiver outputs were present at all times. The observer controlled a switch; when a signal was present, the observer could cause it to appear or disappear from the receiver outputs by turning the switch on or off; when the signal was absent, the position of the switch made no difference in the receiver outputs. The situation thus simulated was one in which the observer could, for purposes of comparison with the background noise, effectively turn the signal on or off by, for example, turning the receiver array.

A summary of the parameters and characteristics of the signal and the noises follows. The signal was Gaussian and stationary, and had a band-pass spectrum with a Q of 10 and a center frequency of 1,000 cps (cycles per second). The unmodulated noises were similar to the signal. They were statistically independent of each other and the signal. The low-frequency noise was Gaussian and stationary, and had a band-pass spectrum with a Q of 1/2 and a center frequency of 7.5 cps; the upper and lower half-power frequencies were 18.1 and 3.1 cps, respectively. The modulating waveform was the above low-frequency noise after strong clipping. The intensities of both noises were simultaneously switched between two levels by the modulator: a high level when the low-frequency noise was positive, and a low level when it was negative. The difference in the noise power at these levels, in decibels (db), will be referred to as the depth of modulation M. The depth of modulation was varied in four steps from 0 to 18.4 db. The average noise power in each of the simulated receiver outputs was always the same; this was also the case for the signal power.

INSTRUMENTATION

The instrumentation will be described briefly. A more complete description has been given previously.[12–14] The audio-frequency noise generators utilized a gas regulator tube as the basic noise source, followed by three stages of amplification. A band-pass filter in the second stage gave the desired spectrum shape, and a cathode follower was used as the final stage to provide a low output impedance.

The low-frequency noise generator employed gamma radiation as a source of short random pulses.[13] Gamma rays excited light pulses in a scintillation crystal which were amplified with a photomultiplier tube. The amplified pulses were then passed through a band-pass filter to give the required spectrum shape. The average rate of occurrence was sufficiently large that the output noise was very nearly Gaussian. Diode clippers were used to obtain the strongly clipped low-frequency modulating voltage.

The modulators were bridge-type modulators with germanium diodes as the rectifying elements. The modulating voltage alternately switched the diodes of the bridge on and off thereby causing the a-c impedance of the modulator to switch between two values; this impedance determined the gain of an amplifier stage which amplified the noise voltage to be modulated. The output of the amplifier was therefore an audio-frequency noise whose intensity switched between two levels determined by the sign of the clipped low-frequency modulating voltage.

The multiplier used in the detector was an analog multiplier of the quarter-difference-squares type.[12] The multiplier output was averaged by a single low-pass *R-C* filter, and the filter output was observed with a d-c oscilloscope having a long-persistence screen.

EFFECTS OF NONSTATIONARY NOISE

As noted previously, a measurement of the average output signal-to-noise ratio as a function of the average input signal-to-noise ratio does not adequately describe the performance of a detector operating with nonstationary inputs. With modulated noise inputs, the expected behavior of the detector employed here would be somewhat as follows: For averaging times short compared with the average period of the modulating waveform (roughly, the reciprocal of its center frequency, or about 0.13 second), the noise output of the averager would increase and decrease in intensity with the modulating wave; after a period of high-level noise output there would follow a low-level noise period since the averager could recover quickly from the high-level input noise. The noise during the low-level periods would determine the threshold levels, since an observer could ignore the high-level noise. Of course, a small signal which existed only during a high-level period could not be detected, but the threshold of a signal occurring during a low-level period or one lasting for several periods would be determined by the low-level noise. With increasing averaging time, the averager would not recover as quickly from the high-level noise, and the threshold level might be expected to increase if the difference between the high-level and low-level noise were sufficiently great. Finally, for averaging times long compared to the modulation period, the averager would not recover at all, and the high-level noise would determine the threshold; the threshold level would then decrease with increasing averaging time.

The Experiment

Experimental Conditions and Procedure

For the detection trials, the averaged multiplier output was continuously displayed on the oscilloscope, with the total sweep time always set at slightly more than ten times the time constant of the averager. Prior to each trial the signal at the multiplier inputs was either turned on or off at random, with a probability of 1/2 for either event. The observer had at his command a switch with two positions: SIGNAL and NO SIGNAL. When the observer's switch was turned to the SIGNAL position the signal appeared at the two multiplier inputs if it had been turned on at the beginning of the trial; otherwise, the signal did not appear. At the start of the trial, the observer's switch was in the NO SIGNAL position. He was allowed to observe the oscilloscope display and turn his switch to SIGNAL once, at any instant of his choosing, then back once to NO SIGNAL, whenever he chose. The switching times and the duration of the observation were not further restricted. At the end of the trial the observer was required to decide whether or not the signal was on; he was then informed of the state of the signal and another trial was made. Although the sweep time was fixed, the observer was allowed to adjust the gain and intensity controls. In the stationary case an rms output-noise deflection of about 0.5 cm (centimeter) was usually satisfactory and the threshold signal was about the same magnitude. For short averaging times and a high level of modulation, however, the threshold was determined by low-level noise and, to see this, the gain was increased until only the low-level noise could be observed. Most of the high-level noise was off the face of the oscilloscope and, effectively, limiting was introduced at the averager output.

For a given set of parameters, the betting curve was obtained in the usual way.[15] The threshold signal level was estimated approximately, and, starting at this level, a series of 40 trials was made; after making allowance for chance successes, the proportion of correct responses was recorded. The signal power was then reduced by 1 db and another series of 40 trials was made. When the 50% point was reached, the signal was increased above the original level in 1-db steps until the 90% point was surpassed. Usually five series of trials served to determine the 90% threshold level to within about 1 db.

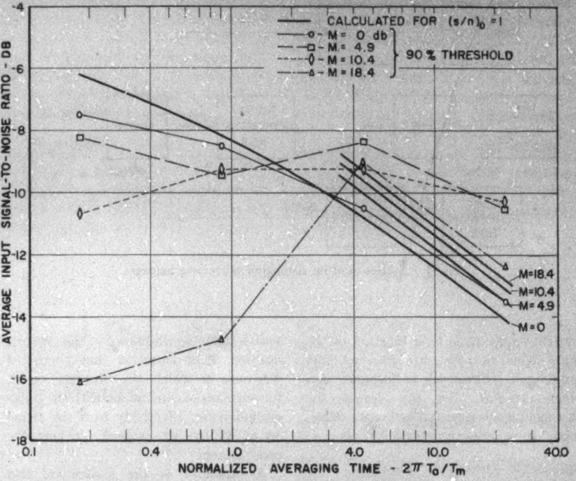


Fig. 2. Experimentally observed performance as function of averaging time; comparison with theoretical performance for long averaging times

Threshold measurements were made for four different modulation levels; starting with no modulation, as a check of the operation of the equipment, the depth of modulation was increased progressively by 4.9, 10.4, and 18.4 db. For each value of M, the 90% threshold level was determined for four averaging times ranging in multiples of 5 from 0.00382 to 0.477 second. Thus the half-power frequency of the averager was 0.334 cps for the largest averaging time and 41.8 cps for the smallest, and measurements were made both above and below the modulation center frequency. The longest averaging time used was determined primarily by drift in the system.

Results and Comparison with Theory

The experimental points are shown in Fig. 2. The results have been plotted as functions of $2\pi T_a/T_m$, where T_a is the averaging time and T_m is the reciprocal of the center frequency of the unclipped modulating noise. This is just the ratio of center frequency of the unclipped noise to the upper half-power frequency of the averager. As noted previously, an increase in the threshold levels as this ratio is increased through unity is to be expected in cases where the modulation depth is large.

In reference 8 it is shown that in the case under study here, where the inputs to the multiplier have equal signal-to-noise ratios and equal noise powers, the output signal-to-noise ratio for long averaging times will be given by equation 47 of reference 8 as:

$$(S/N)_0 = \frac{4\omega_f T_a (S/N)_i^2}{1+(\delta^2-1)^2(\delta^2+1)^{-2}} \qquad (1)$$

provided that the bandwidth of the low-frequency noise prior to clipping is much smaller than the bandwidths of the signal and the unmodulated noises. Here $(S/N)_i$ is the signal-to-noise power ratio at either of the multiplier inputs, $2\omega_f$ is the bandwidth of the signal and of the unmodulated noises in radians per second, T_a is the time constant of the averager, and δ^2 is the ratio of the upper-level noise power to the lower-level noise power. In the case studied here the bandwidth of the unclipped low-frequency noise was 15 cps and that of the signal and unmodulated noises was 100 cps. Since the former was not very small compared to the latter, the assumption made in deriving equation 1 was not strictly true. Accordingly, an analysis has been made taking the foregoing bandwidths into account, and a correction to equation 1 derived. The methods employed parallel closely those followed in reference 8, and the result differs little from equation 1; hence, in view of the length of the derivation, it will not be given here. The result is:

$$(S/N)_0 = \frac{4\omega_f T_a (S/N)_i^2}{1+0.69(\delta^2-1)^2(\delta^2+1)^{-2}} \tag{2}$$

where it has been assumed that $2\omega_f T_a$ is much smaller than unity, i.e., the bandwidth of the averager is much smaller than the bandwidth of the signal and the unmodulated noises. The depth of modulation M is given by:

$$M = 10 \log_{10}(\delta^2) \tag{3}$$

For comparison with the experimental results, the input signal-to-noise ratios necessary for an output signal-to-noise ratio of unity have been determined as functions of $2\pi T_a/T_m$ for modulation depths of 4.9, 10.5, and 18.4 db. These have been plotted as solid lines in Fig. 2. Plots have been given only for the longer averaging times where the assumptions made in deriving equation 2 are valid. In addition, a solid line giving the input signal-to-noise ratio for an output signal-to-noise ratio of unity has been plotted in Fig. 2 for the case where $M=0$. At the two longest averaging times this signal-to-noise ratio differs little from that which would be found from equation 2. Larger differences occur at the shorter averaging times for two reasons. First, at the shortest averaging time, the bandwidth of the averager was comparable to the bandwidth of the signal and the noises; a correction which affects the result only at the shortest averaging time was made to take this into account. Second, at the shortest averaging times, the a-c signal power in the detector output was a definite factor in determining the threshold level; the squared character of the signal in the output was readily apparent and helped in distinguishing the signal. Accordingly the a-c signal power, both that due to the squaring of the signal and that due to multiplication with the noise, at the detector output was counted as signal power in calculating the output signal power. This lowers the required input signal power appreciably at the two shortest averaging times and has negligible effect for the longer times. The experimental points for the stationary case are seen to be in good agreement with theoretical expectations. For the nonstationary cases the experimental results for long averaging times are in general agreement with theory, although the dispersion is greater.

The long-persistence screen employed made the averaging more effective at shorter averaging times since the observer could base his judgment on a large number of after-images, while at longer averaging times only one or two after-images were visible. The possible effect of this is not inconsistent with the results in the stationary case: at short averaging times the experimental points are below the theoretical curve while at longer averaging times they are above it. However, the differences are small and might not be significant.

It was convenient to plot the ratio of the signal power at the 90% threshold level to the average noise power in order to facilitate comparison with theory at long averaging times. Since this average noise power is determined primarily by the power at the upper level, the lowered thresholds for short averaging times shown in Fig. 2 are only a reflection of the fact that the signal was detected in these cases by observations with respect to the noise at the lower level, as was noted previously. If the ratio of the threshold signal power to the noise power at the lower level were plotted it would be found that the points all lie above those for the stationary case. To allow comparison the upper and lower noise levels referred to the average noise power are given in the following table:

Modulation Depth M, Db	High-Level Noise Power, Db	Low-Level Noise Power, Db
4.9	+1.8	− 3.1
10.4	+2.6	− 7.8
18.4	+2.9	−15.5

The general trend of the experimental results is in accordance with that expected. In particular, for the largest depth of modulation the threshold level increases quite rapidly shortly after the ratio $2\pi T_a/T_m$ increases through unity but decreases when this ratio is greater than 5. The same general behavior, though less pronounced, is exhibited at smaller depths of modulation.

It is well known that the threshold signal level for a detector operating with stationary inputs decreases with an increase of averaging time; in marked contrast, the reverse can be true with nonstationary noise as demonstrated here. This behavior results from the fact that at long averaging times it is the power of the high-level noise at the input, not the average input noise power, which determines the noise power at the detector output, and it is clear that some form of limiting might be employed to advantage.

One method of achieving limiting action would be to strongly clip both inputs to the detector prior to multiplication. The multiplier output would then be either of two voltages: a positive voltage when the two inputs were of the same polarity, or a negative voltage, equal in magnitude to the first, when the two inputs were of opposite polarity. The combination of the clippers and the multiplier forms a polarity-coincidence correlator, and the performance of such a device as a detector has been studied both theoretically and experimentally for stationary inputs.[9,10] In addition, theoretical studies of its performance have been made for various types of nonstationary noise.[8] In the cases examined it was found that, although the polarity-coincidence correlator had a slightly poorer performance for stationary noise, its performance was considerably better than that of the multiplier followed by the averager for noise which was markedly nonstationary. On the assumption that the bandwidth of the unclipped low-frequency noise is small compared to the band width of the signal and the unmodulated noise (which was very nearly true in the cases considered here), the ratio of the output signal-to-noise ratio of the polarity-coincidence correlator to that of the multiplier for equal inputs was found to be, as equation 84 of reference 8:

$$R = 0.83\left[1+\frac{1}{4}\left(\delta-\frac{1}{\delta}\right)^2\right]\left[1+\frac{1}{2}\left(\delta-\frac{1}{\delta}\right)^2\right] \tag{4}$$

where δ has been defined previously. For comparison, R in decibels has been found for the cases studied here, and is tabulated as:

Modulation Depth M, Db	R, Db
0.0	−0.8
4.9	+2.9
10.5	11.8
18.4	27.1

The use of the polarity-coincidence correlator can therefore result in a considerable improvement in performance because of the limiting action of this device.

Conclusions

The performance of a detector operating with nonstationary noise has been studied experimentally, and some of the important effects of such noise on performance have been demonstrated here. Furthermore, for long averaging times, the results obtained here have been shown to be consistent with theory. While only one class of nonstationary noise was employed in the measurements, it is reasonable to expect that the behavior of the detection system used will exhibit the same general characteristics for nonstationary noise having more complex low-frequency modulation envelopes. For moderate and high degrees of nonstationarity there will be regions where the threshold signal level will increase with increasing averaging time, and in such cases limiting or clipping prior to detection will result in lowered threshold signal levels.

References

1. Distribution of the Duration of Fades in Radio Transmission—Gaussian Noise Model, **S. O. Rice.** *Bell System Technical Journal*, New York, N. Y., vol. 37, 1958, pp. 581–635.
2. Detection of Fluctuating Pulsed Signals in the Presence of Noise, **P. Swerling.** *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, New York, N. Y., vol. IT-3, 1957, pp. 175–78.
3. Effects of Signal Fluctuation on the Detection of Pulse Signals in Noise, **M. Schwartz.** *Ibid.*, vol. IT-2, 1956, pp. 66–71.
4. Signal-Detection Studies, with Applications, **E. L. Kaplan.** *Bell System Technical Journal*, vol. 34, 1955, pp. 403–37.
5. Underwater Ambient Noise, **V. O. Knudsen, R. S. Alford, J. W. Emling.** *Journal of Marine Research*, New Haven, Conn., vol. 7, 1948, pp. 410–29.
6. Survey of Underwater Sound: Report No. 3, Ambient Noise, **V. O. Knudsen, R. S. Alford, J. W. Emling.** *OSRD Report 4333*, Office of Scientific Research and Development; also *NDRC Report 6.1-1848*, National Defense Research Committee, Washington, D. C., 1944.
7. Arrays for the Investigation of Ambient Noise in the Ocean, **V. C. Anderson.** *Journal*, Acoustical Society of America, New York, N. Y., vol. 30, 1958, pp. 470–77.
8. On the Detection of Signals in Nonstationary Noise by Product Arrays, **J. B. Thomas, T. R. Williams.** *Ibid.*, vol. 31, Apr. 1959 pp. 453–62.
9. Correlators for Signal Reception, **J. J. Faran, Jr., R. Hills, Jr.** *Technical Memo no. 27*, Acoustics Research Laboratory, Harvard University, Cambridge, Mass., 1952.
10. The Application of Correlation Techniques to Acoustic Receiving Arrays, **J. J. Faran, Jr., R. Hills, Jr.** *Technical Memo no. 28*, Acoustics Research Laboratory, Harvard University, 1952.
11. Analysis of a Multiple Receiver Correlation System, **M. J. Jacobson.** *Journal*, Acoustical Society of America, vol. 29, 1957, pp. 1342–45.
12. An Electronic Multiplier, **K. Joetten, T. R. Williams.** *Technical Report no. 10, N6onr-27005*, Department of Electrical Engineering, Princeton University, Princeton, N. J., 1955.
13. A Low-Frequency Random Noise Generator, **S. R. Parris, T. R. Williams.** *Technical Report no. 11, Ibid.*, 1957.
14. Final Report, including "Experimental Studies of Signal Detection in Nonstationary Background Noise," **T. R. Williams.** *N6onr-27005*, Department of Electrical Engineering, Princeton University, 1958.
15. Threshold Signals (book), **J. L. Lawson, G. E. Uhlenbeck.** McGraw-Hill Book Company, Inc., New York, N. Y., 1950, pp. 185–93.

Synthesis of Cable Simulation Networks

ROBERT W. DeMONTE
MEMBER AIEE

Synopsis: This tutorial paper describes a simple design procedure which permits excellent simulation of cable performance employing a symmetrical lattice network containing a small number of elements. The problem is reduced to one of synthesizing 2-terminal impedances at two frequencies. The design technique is illustrated with applications to three lengths of a particular cable. The performance covering more than 2 decades of frequency shows a design error of less than 0.5% in both real and imaginary parts of the image impedance and propagation constant of the cable. With the aid of loading coils, such networks can be used to simulate the transmission characteristics of any type of loaded cable facility having any fractional length of end section.

THE GENERAL DESIRE to improve the transmission performance of exchange area facilities in the Bell System has created needs for precision networks to simulate both image impedance and propagation constant characteristics of nonloaded and loaded cable for test purposes.

The simple design procedure described herein synthesizes these networks in a straightforward manner avoiding any cut and try procedure. This method is based on the open- and short-circuit impedances of the length of cable or on its primary constants at frequencies located at or near the edges of the desired transmission range. From these data, two 2-terminal impedances are synthesized and connected in the form of a symmetrical lattice structure. The crossarm contains 4 elements or less and the series arm 3 elements or less. Both its image or characteristic impedance and propagation constant very closely approximate that of the physical cable over the entire transmission range of frequency, which ranges from about 50 cps (cycles per second) to 10 kc.

When compared with other design procedures, this method appears superior for specified number of elements with respect to degree of simulation for given length of frequency.

Theory

Background

The classical theory on the design of simulating networks relating to uniform telephone cables has been developed by Bartlett.[1] Guillemin[2] has extended the theory by assuming a fixed number of terms used in the infinite product expansion of tanh P and proceeds to determine the deviation for a given frequency coverage. These methods appear to have little practical advantage since they assume that the primary constants do not vary with frequency and also because they require many elements in the resultant network. Johnson[3] has shown a cut and try method employing various arrangements of ladder-type networks and models of such designs have been in laboratory use for more than 25 years. When originally constructed the design requirements involved a rather narrow frequency band and adequate data on primary constants were lacking.

Zobel[4] has considered the same general problem for a length equivalent to about 10,000 feet and has a double section lattice structure containing a type of m derived 2-terminal impedance which is related directly to the cable constants (cable constants are assumed to be independent of frequency). While such a method is expected to simulate the cable performance to a high degree, it appears to require more elements than are needed by the method described herein. He has also shown an elegant method of synthesizing single constant R lattice sections to simulate the propagation constant using only the attenuation constant of the cable at two frequencies.

Paper 59-774, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 4, 1959; made available for printing April 8, 1959.

Robert W. DeMonte is with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

The writer wishes to express his appreciation to W. R. Lundry for his inspiring discussions, to Miss J. D. Goeltz for computer programing, and to Lester Hochgraf for suggesting the problem.

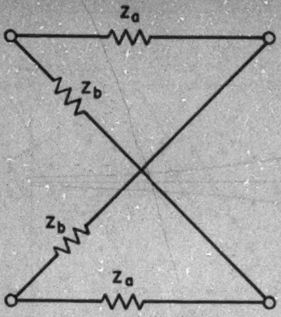


Fig. 1. Lattice representation of cable

LATTICE EQUIVALENT OF CABLE

A cable having primary constant of R, L, G, and C per unit length and length l is characterized by an image impedance

$$z_k = \sqrt{(R+jL\omega)/(G+jC\omega)} = r_k - jx_k \tag{1}$$

and the propagation constant

$$Pl = l\sqrt{(R+jL\omega)(G+jC\omega)} = \alpha + j\beta \tag{2}$$

where

$\omega = 2\pi f$ (f = frequency)
α = the attenuation constant
β = the phase shift constant

In terms of the open- and short-circuit impedance measurements z_{oc} and z_{sc}, respectively, the image impedance

$$z_k = \sqrt{z_{oc} z_{sc}} = r_k - jx_k \text{ (ohms)} \tag{3}$$

and the propagation constant

$$Pl = 10 \log_{10} \left|\frac{1+\sqrt{z_{sc}/z_{oc}}}{1-\sqrt{z_{sc}/z_{oc}}}\right| + \text{angle of} \left|\frac{1+\sqrt{z_{sc}/z_{oc}}}{1-\sqrt{z_{sc}/z_{oc}}}\right| \tag{4}$$

$$= \alpha(db) + j\beta(\text{deg.})$$

It is well known that any distributed system of length l such as a cable can be represented by a symmetrical lattice structure as shown in Fig. 1, where

$$z_a = z_{oc}\left(1-\sqrt{1-\frac{z_{sc}}{z_{oc}}}\right) = z_k \frac{\epsilon^{Pl}-1}{\epsilon^{Pl}+1} \tag{5}$$

and

$$z_b = z_{oc}\left(1+\sqrt{1-\frac{z_{sc}}{z_{oc}}}\right) = z_k \frac{\epsilon^{Pl}+1}{\epsilon^{Pl}-1} \tag{6}$$

are the series and crossarm impedances respectiely. They are also equal to the short- and open-circuit impedances of a length $l/2$.

Then

$$z_n = \sqrt{z_{ab} z_b} = r_n - \mathrm{J}x_n \tag{7}$$

and

$$Pl_n = 20 \log_{10} \left|\frac{1+\sqrt{z_a/z_b}}{1-\sqrt{z_a/z_b}}\right| + \text{angle of} \left|\frac{1+\sqrt{z_a/z_b}}{1-\sqrt{z_a/z_b}}\right| \tag{8}$$

$$= \alpha_n + j\beta_n$$

are the image impedance and propagation constants of the equivalent lattice network respectively.

Now the problem is one of determining the configuration of the network and element values for the two two-terminal impedances to approximate z_a and z_b over the desired frequency range.

METHOD OF DESIGN

Direct computations show that for short lengths of cable of 6,000 feet or less, both z_a and z_b are modestly increasing or decreasing monotonic curves over the frequency range up through 10,000 cps as shown in Figs. 2 and 3. f_1 and f_2 represent frequencies at or near the ends of the transmission band. In Fig. 2, for a cable length l_1 the curves marked r_1 and x_1 are the impedance components of z_a when it has inductive reactance. As the length is increased to l_2, that rare case is obtained when the reactive component x_2 approximates zero over the frequency range and is zero at f_2. r_2 closely approximates a constant resistance which is nearly equal to one-half of the d-c resistance of the cable. Greater lengths are represented by curves r_3 and x_3 which indicate a capacitive reactance component of z_a.

In Fig. 3, r_b and x_b are typical shapes of the components of z_b for all lengths. In the vicinity of frequency f_2, the reactance curve x_b tends to approach the zero value more rapidly for the longer lengths of cable.

The primary constants of a real cable vary with frequency. Skin effect increases the resistance and decreases the inductance. The conductance increases several hundred fold while the capacitance decreases a small amount. These variations are accounted for in the shape of the curves of Figs. 2 and 3.

Because of the simple shapes of these empirically determined graphs, when synthesized at two frequencies say f_1 and f_2, the approximant is accurate at all frequencies within the band as well as in the vicinity of its edges.

A. SYNTHESIS OF z_a

Each pair of impedance curves in Fig. 2 can be represented by a simple ratio of polynominals.

Let

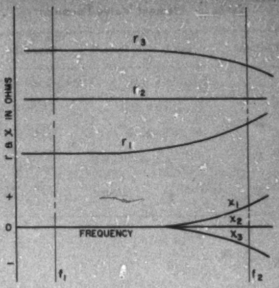


Fig. 2. Impedance of series arm z_a

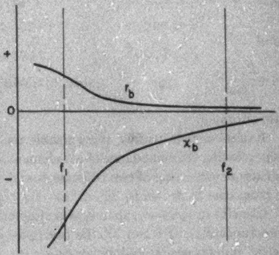


Fig. 3. Impedance of crossarm z_b

$$z_a = R_a + jx_a = \frac{a_0 + a_1 p}{1 + b_1 p} \tag{9}$$

where $p = j\omega$

The coefficients a_0, a_1 and b_1 are determined as follows:

At $\omega = \omega_1$

$$z_1 = R_1 + jx_1 = \frac{a_0 + ja_1\omega_1}{1 + jb_1\omega_1} \tag{10}$$

since

$x_1 \to 0$ as $\omega_1 \to 0$

$$a_0 = R_1 \tag{11}$$

At $\omega = \omega_2$

$$z_2 = R_2 + jx_2 = \frac{a_0 + ja_1\omega_2}{1 + jb_1\omega_2} \tag{12}$$

which is satisfied when

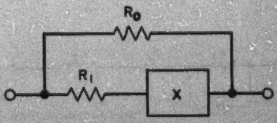


Fig. 4. General network of series arm z_a

Table I. Element Value Formulas

Element	Case I	Case II	Case III
	$b_1>\frac{a_1}{a_o}$	$b_1=\frac{a_1}{a_o}$	$b_1<\frac{a_1}{a_o}$
R_o	$\frac{a_1}{b_1}$	a_o	a_o
R_1	$\frac{a_oa_1}{a_1-a_ob_1}$	infinite	$\frac{a_1a_o}{b_1a_o-a_1}$
X	$L=\frac{a_1^2}{a_1+a_ob_1}$	infinite	$C=\frac{a_ob_1-a_1}{a_o^2}$

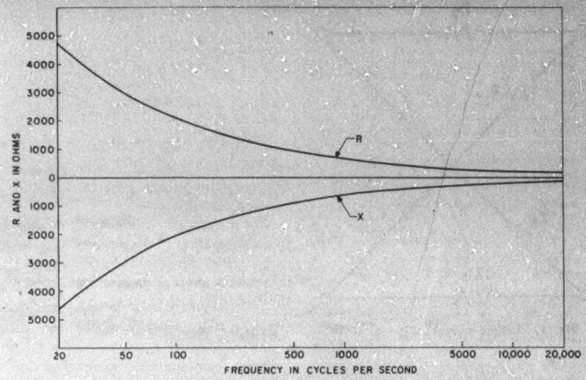


Fig. 6. Impedance of cable

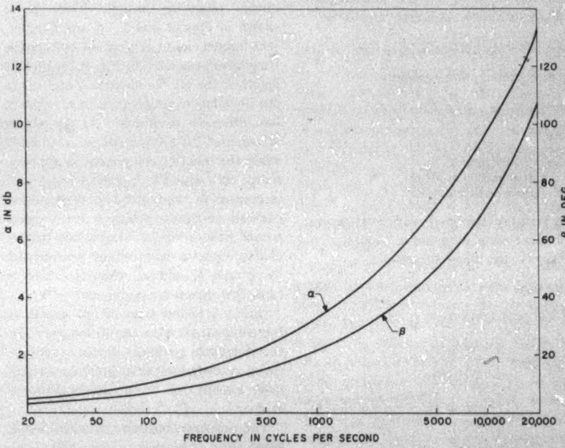


Fig. 7. Propagation constant of cable

$R_2-b_1x_2\omega_2=a_o$ and

$$x_2+b_1R_2\omega_2=a_1\omega_2 \tag{13}$$

Solving equations 13 for a_1 and b_1,

$$a_1=\frac{x_2^2+R_2(R_2-a_o)}{x_2\omega_2} \tag{14}$$

and

$$b_1=\frac{R_2-a_o}{x_2\omega_2} \tag{15}$$

are obtained.

Equation 9 is readily recognizable as representing the impedance of a 3-element structure consisting of two resistances and a reactance as shown in Fig. 4.

Clearly, if $a_1/a_o=b_1$, then z_2 is a simple resistance R_o. If $b_1>a_1/a_o$, the reactance is an inductance L and if $b_1<a_1/a_o$, the reactance is a capacitance C.

The element values may be computed from the formulas in Table I.

In Case III when $a_1=0$, $R_1=0$, $C=b_1/a_o$, and $R_o=a_o$.

In Case II $z_a=r=Rl/2$ where R is the d-c resistance per unit of length of cable. The condition which must exist for $x_a=0$ will now be developed.

Equation 5 can be written as

$$z_a=z_k\frac{\epsilon^{Pl}-1}{\epsilon^{Pl}+1}=z_k(m+jn) \tag{16}$$

From equations 1 and 16

$$z_a=(r_k-jx_k)(m+jn)=(r_km+x_kn)+j(r_kn-x_km) \tag{17}$$

In equation 17 the imaginary part becomes zero when

$$r_kn-x_km=0$$

$$\frac{x_k}{r_k}=\frac{n}{m} \tag{18}$$

then equation 17 becomes

$$z_a=r_km+x_kn=\frac{Rl}{2} \tag{19}$$

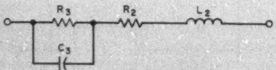


Fig. 5. General network of crossarm z_b

Since it is easy to show that $(m+jn)$ is proportional to (sinh $\alpha l+j$ sin βl), this equation holds when

$$\frac{\sin l\beta}{\sinh l\alpha}=\frac{x_k}{r_k} \tag{20}$$

B. SYNTHESIS OF z_b

Referring to Fig. 3, the impedance of the crossarm z_b is obviously composed of two parts; a resistance in parallel with a capacitor describing the low-frequency region plus a resistance and a small inductance connected in series to correct for the reactance at the higher frequencies.

The complete structure for z_b is shown in such as Fig. 5. Referring to the graphs in graphs in Fig. 3, at some low frequency such as f_1, the impedance

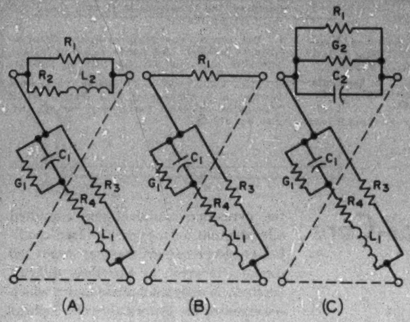
$z_b=r_1-jx_1$ and its admittance is (21)

$$Y_b=G+jB=\frac{r_1}{r_1^2+x_1^2}+j\frac{x_1}{r_1^2+x_1^2} \tag{22}$$

Now

$$R_3=\frac{1}{G}=\frac{r_1^2+x_1^2}{r_1} \tag{23A}$$

and



SYMBOL	(A)	(B)	(C)
R_1	167.36 OHMS	177.2 OHMS	250.0 OHMS
R_2	493.0 OHMS	—	—
R_3	34.7×10^6 OHMS	$14. \times 10^6$ OHMS	14.3×10^6 OHMS
R_4	46.0 OHMS	62.0 OHMS	86.2 OHMS
C_1	0.02245 μF	0.03172 μF	0.04484 μF
G_1	1.4×10^{-8} $+ 2 \times 10^{-11}$ FREQ	1.9×10^{-8} $+ 3.72 \times 10^{-11}$ FREQ	2.8×10^{-8} $+ 4 \times 10^{-11}$ FREQ (MHOS)
C_2	—	—	0.00752 μF
G_2	—	—	0.47×10^{-8} $+ 0.67 \times 10^{-11}$ FREQ. (MHOS)
L_1	0	0	0.00007 h
L_2	0.00910 h	—	—
f_1	50 CPS	50 CPS	50 CPS
f_2	10,000 CPS	10,000 CPS	10,000 CPS
EQ LENGTH	3000 FT	4229 FT	6000 FT

Fig. 8. Schematics and element values of networks simulating three lengths of cable

$$C_3 = \frac{B}{\omega_1} = \frac{x_1}{\omega_1(r_1^2 + x_1^2)} \tag{23B}$$

The resistance R_3 is generally more than the reactance of C_3 at f_1 so that at the higher frequency f_2, which may be 100 f_1, its effect is negligible. At f_2,

$$z_b = r_2 - jx_2 \tag{24}$$

so a resistance of value r_2 needs to be added in series. Now, x_2 is smaller than the reactance x_c of C_3 at f_2, and the difference must be made up by an inductance, that is

$$\frac{x_c - x_2}{\omega_2} = L_2 \tag{25}$$

when has a value of the order of a millihenry. x_2 is less than x_c when the distributed capacitance of the cable is fairly constant and decreases slightly with frequency and when the lengh of the cable is not too small.

To complete the design of the lattice structure, the conductance associated with physical capacitors should be included to obtain the correct performance of the network. If the silvered mica type[5] of capacitors are employed the effect of conductance is small in the frequency range up to 10,000 cps.

The complete schematics of the simulating networks for three cases are as shown in Fig. 8. The high resistance in the crossarm has been connected across the entire branch for convenience and the effect of this change is negligible throughout the frequency range.

Applications

Based on the preceding methods and on data represented by the curves of image impedance and propagation constants for a particular cable shown in Figs. 6 and 7 respectively, three simulating networks have been designed which are equivalent to lengths of 3,000, 4,229, and 6,000 feet. Network schematics with element values and the control frequencies are shown in Fig. 8. The conductances G_1 and G_2 are the inherent conductances of the capacitors C_1 and C_2 respectively.

Calculations were made on the performance of the networks and their deviations or design errors have been plotted in Figs. 9 and 10. In Fig. 9 are shown the reflection coefficients

$$\rho = \left|\frac{z_k - z_{in}}{z_k + z_{in}}\right|$$

in per cent versus frequency where z_{in}

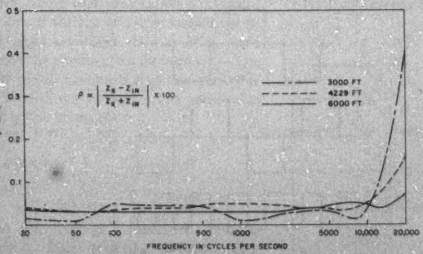


Fig. 9. Reflection coefficient: network versus cable

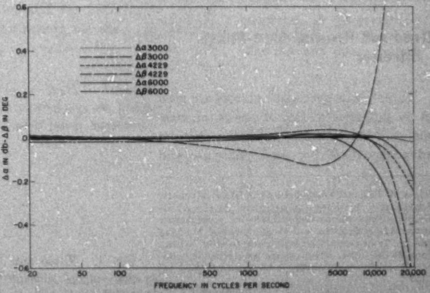


Fig. 10. Design error of propagation constants

is the input impedance of the network terminated at the far end in z_k.

In Fig. 10 are shown the design errors in both components of the propagation constants

$\Delta\alpha = \alpha_n - \alpha$ in decibels
$\Delta\beta = \beta_n - \beta$ in degrees

References

1. The Theory of Electrical Artificial Lines and Filters, A. C. Bartlett. John Wiley & Sons, Inc., New York, N. Y., 1931, chap. V, pp. 75–84.

2. Communication Networks, E. A. Guillemin. John Wiley & Sons, Inc., 1935, chap. VII, pp. 250–79.

3. Transmission Circuits for Telephonic Communication, K. S. Johnson. D. Van Nostrand Company, Inc., Princeton, N. J., 1925, chaps. XII and XIII, pp. 143–60.

4. Distortion Correction in Electrical Circuits, O. J. Zobel. *Bell System Technical Journal*, New York, N. Y., vol. VII, July 1928, pp. 438–534.

5. Mica Capacitors for Carrier Telephone Systems, A. J. Christopher, J. A. Kater. *AIEE Transactions (Electrical Engineering)*, vol. 65, Oct. 1946, pp. 670–74.

A More Stable 3-Phase Transistor-Core Power Inverter

W. E. JEWETT
ASSOCIATE MEMBER AIEE

P. L. SCHMIDT
ASSOCIATE MEMBER AIEE

THE 3-phase transistor-core power inverter described in this paper follows the trend away from mechanical devices toward lighter, more efficient, electronic equipment which, in most cases, are more reliable than the mechanical devices they replace. A single-phase transistor-core inverter has already replaced the mechanical multivibrator in one of our missile systems; it is only natural, then, for an electronic 3-phase inverter to follow as a power source for the 3-phase gyroscope motors which are part of the instrumentation in the same missile. This paper presents:

1. The need for a 3-phase inverter and discusses the reasons for the use of 3-phase gyroscope motors rather than d-c single-phase or 2-phase motors.
2. The waveshape characteristics of a 3-phase square-wave system and points out some of the system's limitations.
3. The operation of the 3-phase inverter which is an extended version of the one described by A. G. Milnes.[1]
4. The reasons for adding additional circuitry to Milnes' inverter to make it more stable over wide load ranges.

Need and Reasons for 3-Phase Inverter

Three-phase gyroscope motors are used in the guidance section of one of our missile systems because they offer distinct advantages over d-c, single-phase a-c, and 2-phase a-c motors. The d-c motors have commutators which get dirty and short, brushes which wear and sometimes make poor contact, and mechanical contacts which constantly make and break during the process of speed regulation. All of these mechanical features are highly undesirable from a reliability point of view. Single-phase a-c motors, in addition to coming up to rated speed rather slowly, require starting circuitry. As far as 2-phase motors are concerned, the gyroscope manufacturers consulted all agree that for the specified load conditions and required starting time (30 seconds) it would be extremely difficult for them to design a 2-phase motor which could match the performance of a 3-phase motor of comparable size and weight. They strongly recommended 3-phase motors. The reasons, then, for the choice of 3-phase motors are reliability and performance.

Once the 3-phase motors were decided upon, a static method of powering the motors had to be devised. This meant inverter-type circuitry which generates squarewaves. The question to be answered was: are the waveshape characteristics of a 3-phase square-wave system compatible with available motors which heretofore have been powered by a sine-wave system? The first logical thing to do was to look at these characteristics.

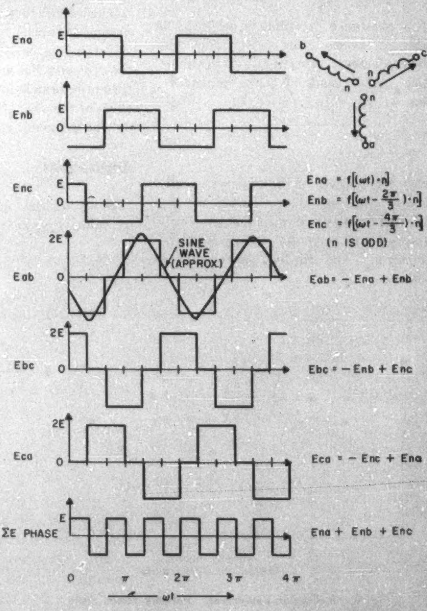


Fig. 1. Phase and line-voltage characteristics of 3-phase square-wave system

---

Paper 59-786, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted November 14, 1958; made available for printing April 14, 1959.

W. E. Jewett and P. L. Schmidt are with Bell Telephone Laboratories, Inc., Whippany, N. J.

Vol.
Pt.
195

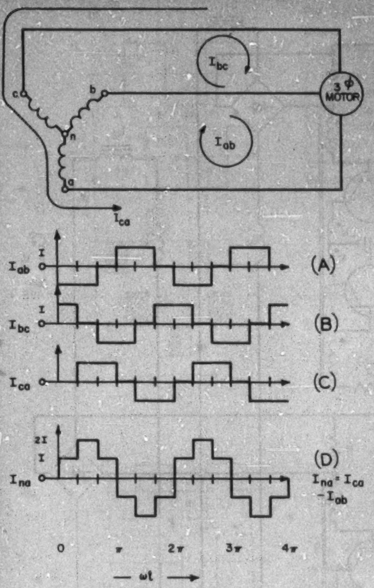


Fig. 2 (left). Current waveshapes in a 3-phase square-wave system with balanced resistive load

A, B, C—Load phase currents when load is delta connected D—One load phase current when load is Y connected

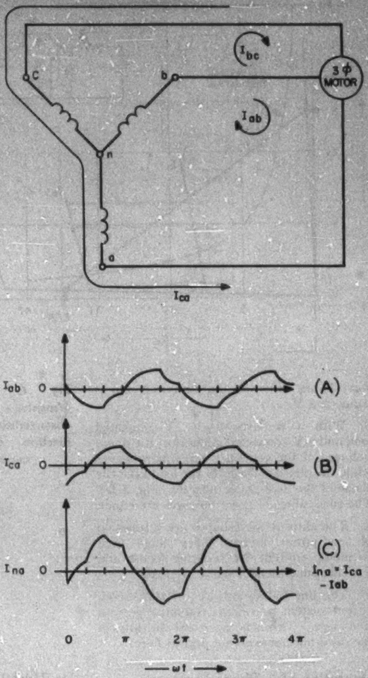


Fig. 3 (right). Current waveshapes in a 3-phase square-wave system with balanced inductive load

A,B—Two load phase currents when load is delta connected C—One load phase current when load is Y connected

Waveshape Characteristics of 3-Phase Square-Wave System

Shown in Fig. 1 are the three phase voltages locked 120 degrees apart with respect to each other, the three line-to-line voltages obtained from the Y connection, and the third harmonic voltage which would cause circulating currents in the delta connection.

In a 3-phase system there are three possible generator connections: 1. Y with neutral; 2. Y with no neutral; and 3. delta.

Since the gyroscope motors are 3-wire there is no need to consider the first. The delta connection is prohibited by the fact that the third harmonic and its odd harmonics in each of the three phases add to give a voltage equal in magnitude to the phase voltage. If the delta were closed the windings would burn up. The only possible connection of this generator is Y with no neutral. The rms value of the line voltage is

$$E_{\text{rms}}\text{ line}=\sqrt{\frac{1}{\pi}\int_{0}^{2\pi/3}(2E_{\text{phase}})^{2}\,d\theta}$$
$$=1.63E_{\text{phase}} \quad (1)$$

The line voltage contains the fundamental, fifth, seventh, eleventh, thirteenth, etc., harmonics. The third and all its odd harmonics add out in the Y connection. A sine wave has been sketched through one of the line voltages showing the resemblance between the two. Thus, the motors are to be driven with a voltage other than a square wave. Actually, the waveshape of the phase voltage the motors experience depends upon whether they themselves are connected in the Y or delta configuration.

Figs. 2(A), (B), and (C) show the three phase currents of the motors if they are internally delta connected, and Fig. 2(D) shows one of the phase currents for the internal Y connection. Actually, Fig. 2 is for a resistive load. Figs. 3(A) and (B) show two of the phase currents for the internal delta connection and Fig. 3(C) shows one of the phase currents for the Y connection of the motors assuming they are inductive, which they are. For a 3-phase square-wave system, conclude the following:

1. The delta connection of the generator is not possible. This is somewhat disadvantageous as only one value of line voltage may be obtained.
2. For the Y connection of the generator, the line voltage and phase voltage waveshapes differ. The line and phase voltage

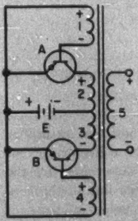


Fig. 4 (left). Schematic of basic d-c to a-c inverter using p-n-p transistors

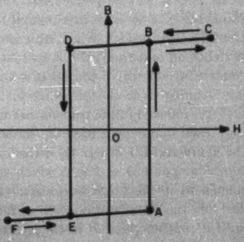
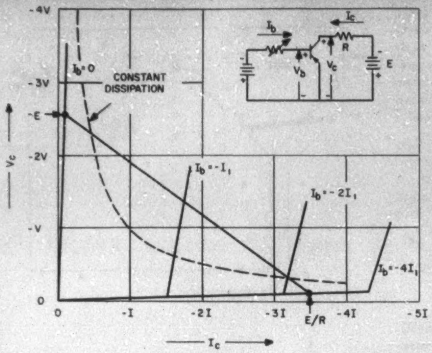


Fig. 5 (right). Typical hysteresis loop of inverter transformer core



waveshapes differ at the load if it too is connected Y.

3. With a resistive 3-wire Y-connected load and a Y-connected generator, the phase voltages of the generator are square waves while the phase voltages of the load take the shape of the load-phase currents, Fig. 2(D). The rms values of these voltages are equal.

4. The current waveshapes are a function of load-power factor. They have high harmonic content for resistive loads, and they are smoother for inductive loads.

5. The line-to-line voltages of the Y-connected system bear some resemblance to a sine wave. This is because odd multiples of the third harmonic do not appear here.

Fig. 6 (above). Transistor - collector characteristics as a function of base current

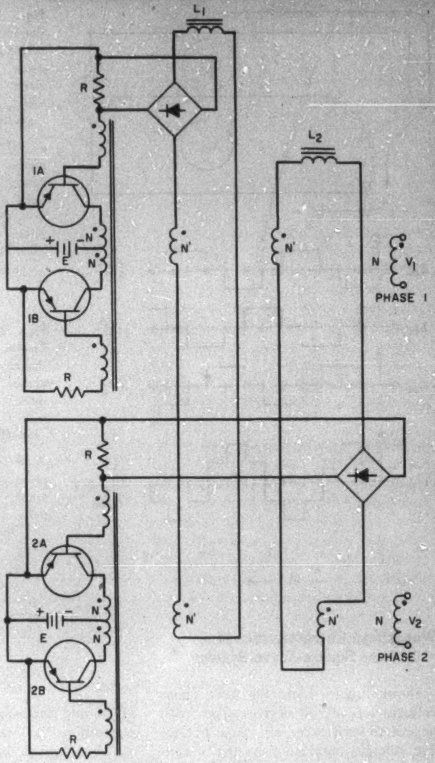


Fig. 7 (right). Schematic of a basic 2-phase d-c to a-c inverter using p-n-p transistors

Operation of 3-Phase Inverter

The inverter consists of three distinct parts: 1. three single-phase d-c to a-c inverters; 2. circuitry to synchronize these single-phase inverters; and 3. three power amplifiers which are driven in push-pull by the single-phase inverters. Each of these will be discussed in the order listed.

BASIC D-C TO A-C SINGLE-PHASE INVERTER

Because the basic inverter circuit has been widely discussed in the literature by Royer,[2] Uchrin and Taylor,[3] Van Allen,[4] and others, only a qualitative description of its action will be given here. Fig. 4 shows the inverter in its most simple form. Consider the instant when transistor *A* switches on and transistor *B* switches off connecting winding *2* to and disconnecting winding *3* from the battery. The voltages induced in the transformer windings then have polarities as shown in the figure and the flux in the transformer core is at point *A* of Fig. 5 which corresponds to its negative saturation value. The flux then increases linearly with time until it reaches point *B* of the hysteresis curve at which time the transformer core saturates in the positive direction. The time require for the transformer core to go from negative to positive saturation is derived from Faraday's law as:

$$t=\int_{-\phi\text{sat}}^{+\phi\text{sat}}\frac{N}{E}\,d\phi=\frac{2N\phi\text{ sat}}{E} \tag{2}$$

giving a frequency of

$$f=\frac{1}{2t}=\frac{E}{4N\phi\text{ sat}} \tag{3}$$

where

E = input voltage (Fig. 4)
N = number of primary turns either side of center tap
ϕ sat = saturation flux

When the transformer core saturates, point *B* of Fig. 5, the collector current increases to a value determined by the base current; thus, the flux in the core increases beyond its value at point *B* to some value at *C* which is slightly larger. At point *C* there is no further change in flux so all the induced voltages in the transformer windings fall to zero. The collector current now begins to fall to zero inducing voltages in the base windings *1* and *4*, which begin to drive transistor *B* into and transistor *A* out of conduction. The flux traverses the path from *C* to *D* quite rapidly and upon reaching point *D* the full battery voltage less some small transistor drop is applied to winding *3* of the transformer. The half cycle then repeats with transistor *B* conducting and transistor *A* nonconducting.

For the single-phase inverter to operate efficiently, the corners of the hysteresis loop must be quite sharp, and the base current only slightly higher than the value required for good output waveshape. Fig. 6 shows the transistor dissipation to be high in the switching region.

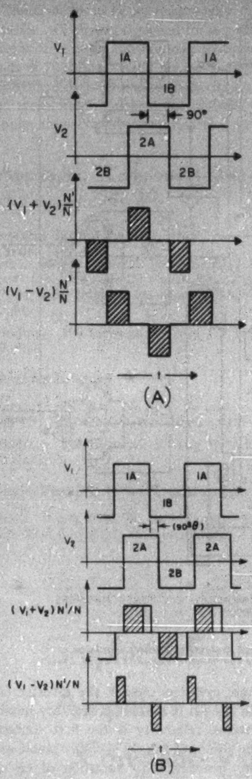


Fig. 8 (left). Waveshapes of circuit of Fig. 7

A—Phases properly locked with V_2 lagging V_1 by 90 degrees

B—Circuit not properly locked; sum circuit takes corrective action

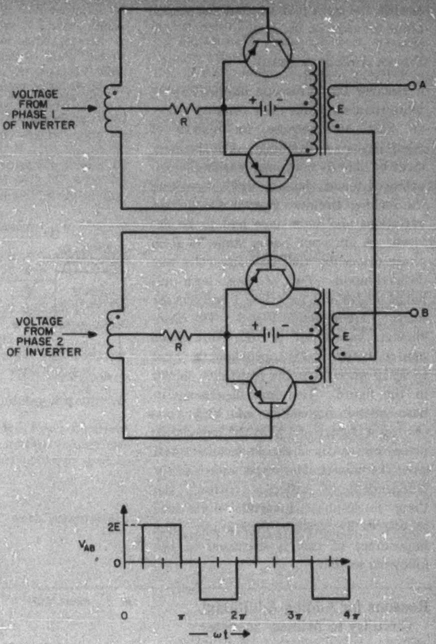

VOLTAGE FROM PHASE 1 OF INVERTER
R
E
A
VOLTAGE FROM PHASE 2 OF INVERTER
R
E
B
2E
V_{AB}
0
0
π
2π
3π
4π
ωt

Fig. 9 (right). Two phases of power amplifier stage, and waveshape characteristic of line voltage

Excessive base current sends the transistor into a region where the dissipation is higher than it need be, and a core material having a hysteresis loop that is not sharp at the corners will keep the transistor in this region longer than is necessary.

SYNCHRONIZING CIRCUITRY

The synchronizing circuit which was described by Milnes is ingeniously simple. It employs two saturating reactors and eight diodes for each pair of inverters to be phase locked. Thus, for a 3-phase system, two reactors and eight diodes are required to lock phase 2 on phase 1, and two reactors and eight diodes are required to lock phase 3 on phase 2. To lock phase 2 on phase 1, circuits which measure the sum and difference voltages of these two phases are used. If the sum and difference voltages are other than some prescribed value which is determined by the phase angle desired, corrective action takes place. Consider the 2-phase circuit shown in Fig. 7. Phase 2 is locked 90 degrees behind phase 1 [Fig. 8(A)]. To accomplish this, L_1 and L_2, the two saturating reactors, must saturate in $T/4$ seconds, where T is the period of the fundamental phase frequency. The sum and difference voltages, (V_1+V_2) N'/N and (V_1-V_2) N'/N respectively, are positive for one-quarter cycle, zero for the next quarter cycle and then repeat negatively. As long as the circuit maintains this 90-degree phase relationship, reactors L_1 and L_2 will completely absorb the sum and difference voltages. Now, suppose for some reason phase 2 is caused to lag phase 1 by some angle less then 90 degrees, say 90 degrees$-\theta$ [Fig. 8(B)]. The sum voltage will be positive for $(T/4+\theta/\omega)$ seconds, where ω is the fundamental angular velocity of the phase voltages. The reactor L_1 saturates in $T/4$ seconds as the magnitude of the voltage applied to it remains unchanged. The voltage (V_1+V_2) N'/N is then applied across resistor R in the base circuit of transistor $1A$ in such a direction as to make the base of the transistor positive with respect to its emitter. Since the transistors are p-n-p types, $1A$ will cut off and $1B$ will conduct. At this instant phase 2 lags phase 1 by 90-degrees again. Corrective action is not complete, however, as the transformer core of phase 1 has undergone only a partial flux swing. Corrective action will continue to take place until the 90-degree relationship with phase 2 lagging is maintained; about six cycles of time are required.

It is obvious from Fig. 8(A) that the sum circuit does the correcting for phase angles between -90 degrees and $+90$ degrees; and the difference circuit does the correcting for angles greater than $+90$ degrees and less than -90 degrees. Phase angles are considered positive when phase 2 lags phase 1.

For phase 2 to lock 120 degrees behind phase 1 requires that L_1 saturate in $T/6$ seconds and L_2 saturate in $T/3$ seconds where T is once again the period of fundamental frequency of the phase voltages. If a third phase is added and locked 120 degrees behind phase 2 a 3-phase system is obtained. The third phase is locked on the second phase in exactly the same manner that the second phase is locked on the first.

POWER AMPLIFIER STAGE

Because the gyroscope motors require more than six times their running power on start, and because, for reasons of reliability and flexibility, the inverter must be able to work into reasonably unbalanced loads, some method of making the locking circuitry insensitive to load variations and unbalance had to be devised. A push-pull power stage for each phase driven by the basic inverters was decided upon. Two of these push-pull power stages, one for phase 1 and one for phase 2, are shown in Fig. 9. The base-emitter impedance of the transistors used in these stages is approximately equal to $R/10$ where R is the resistance shown in the figure. Therefore, a change in base-emitter impedance of 10% constitutes a change of 1% in the impedance presented to the basic inverters. And since the base-emitter impedance is fairly independent of collector current, the three single-phase inverters work into an essentially constant impedance. The importance of this is discussed in the following section.

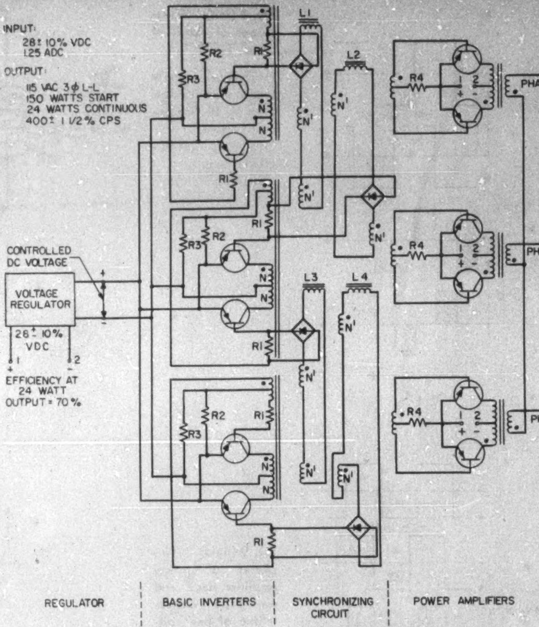


Fig. 10. Complete schematic of 3-phase inverter including regulator

Reasons for Adding Additional Circuitry to Milnes' Inverter

Shown in Fig. 10 is a complete schematic of the 3-phase inverter. The voltage regulator shown as a block diagram is a series-transistor type and is used to control the system frequency to $1^{1}/_{2}\%$ of its nominal 400 cps (cycles per second). This is accomplished by controlling the input voltage to the basic inverter stages for battery voltage changes of $\pm 10\%$ over a temperature range of -40 C (degrees centigrade) to $+65$ C. At constant temperature, the regulator is a constant voltage device within this tolerance of $\pm 1^{1}/_{2}\%$. For varying temperatures the output voltage of the regulator varies in such a manner as to maintain the system frequency to within $\pm 1^{1}/_{2}\%$ of its nominal value. As the basic inverter stages experience an essentially constant impedance, their frequency is a function of input voltage and temperature only.

If the performance of an inverter without the power-amplifier stage is considered, note the following:

1. A simple voltage regulator will no longer regulate the frequency. The frequency will vary when the induced voltages in the basic inverter windings vary and this will happen as the load is varied even if the input voltage is held constant.
2. A discriminator-type regulator must be used to regulate the frequency. It will also be noted that full power must now come through the regulator. This feature is highly undesirable from an efficiency and a size and weight point of view. It does, however, provide a considerable amount of load-voltage regulation which would be desirable for certain types of loads such as filaments. The gyroscope motors performed well for load voltage variations of $\pm 10\%$ so no consideration was given to voltage regulation.
3. For unbalanced loads the locking angle is affected because the induced voltages in the three phases differ. If the unbalance is too great the phases will not lock.
4. The currents which flow in the primary windings of the three inverters will look like I_{na} of Fig. 2 for resistive loads, and I_{na} of Fig. 3 for inductive loads. These currents of odd waveshape affect the voltages applied to the saturating reactors, especially when the motors are started; under this condition these currents are more than six times the value they attain when the motors are running at rated speed.

Our experience in the laboratory has been: for extreme conditions of load variation and load unbalance, a frequency-regulated, more stable, 3-phase transistor-core power inverter could be made by adding a stage of power amplifiers and regulating the voltage into the three single phase inverters. Since the inverter circuit is used in one of our military missile systems, reliability is the first concern. The circuit of Fig. 10 is highly stable and very insensitive to variation of circuit parameters. Even without the regulator, no change in frequency could be detected for a load swing from 24 watts to 150 watts at constant temperature and input voltage.

If two of the phases are loaded for 50 watts each and the third phase unloaded, the only measurable effect on the system is an unbalance of the line-to-line output voltages. This unbalance is of course due to the IR drops in the windings of the loaded phases.

Conclusions

A 3-phase transistor-core power inverter that is almost completely insensitive to the degree of load variation and unbalance has been developed as a power source for 3-phase gyroscope motors in one of our missile systems. Laboratory tests have shown that 3-phase gyroscope motors perform as well when powered by this square-wave inverter as they do when powered from a sine-wave system.

Vol.
Pt. 1
1959

There seems to be no reason why this inverter, which was designed for a power output of 24 watts continuously and peak loads of 150 watts, could not, with slight adjustment of the transistor-base circuitry and transformer-copper area, be made to supply almost 1,000 watts continuously.

Appendix. Component Data, Fig. 10

Reactors L_1 and L_3

Cores: 2 mils 50% Ni (nickel), 50% Fe (iron) tape, 0.5-inch ID (inside diameter), 0.75-inch OD (outside diameter), 0.125-inch height.

Winding: 103 turns no. *30* Awg HF wire, 0.9 ohm.

Reactors L_2 and L_4

Cores: 2 mils 50% Ni, 50% Fe tape, 0.5-inch ID, 0.75-inch OD, 0.125-inch height.

Winding: 206 turns no. *30* Awg HF wire, 1.8 ohms.

Single-Phase Inverter Transformers

Cores: 2 mils 50% Ni, 50% Fe tape, 1.0-inch ID, 1.5-inch OD, 0.25-inch height.

Windings:
1. Primary; 258 turns no. *33* Awg HF wire, 7.7 ohms, either side of center tap.
2. Feedback to base circuit; 21 turns no. *33* Awg HF wire, 0.7 ohm, either side of center tap.
3. Synchronizing circuit; 52 turns no. *33* Awg HF wire, 1.8 ohms.
4. Power amplifier driver; 65 turns no. *33* Awg HF wire, 2.4 ohms, either side of center tap.

Power-Amplifier Transformers

Cores: 4 mils grain oriented silicon steel tape, 1.0-inch ID, 1.625-inch OD, 0.438-inch height.

Windings:
1. Primary; 176 turns no. *26* Awg HF wire, 1.4 ohms, either side of center tap.
2. Secondary; 478 turns no. *28* Awg HF wire, 7 ohms.

Rectifiers in Synchronizing Circuit

100 milliamperes average forward-current rating, 200 volts peak inverse voltage rating, diffused junction silicon or equivalent.

Resistors

*R*1, 10±5% ohms, 1/2 watt, composition, see reference 5.

*R*2, 24±5% ohms, 1/2 watt, composition, see reference 5.

*R*3, 27,000±5% ohms, 1/2 watt, composition, see reference 5.

*R*4, 27±5% ohms, 2 watts, composition, see reference 5.

Transistors

Basic inverter stage, Western Electric type *2N463*, germanium, p-n-p or equivalent.

Power amplifier stage, Delco type *2N174*, germanium, p-n-p or equivalent.

References

1. Phase Locking of Switching-Transistor Converters for Polyphase Power Supplies, A. G. Milnes. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, Nov. 1955, pp. 587–92.

2. A Switching Transistor D-C to A-C Converter Having an Output Frequency Proportional to the D-C Input Voltage, G. H. Royer. *Ibid.*, July, pp. 322–26.

3. A New Self-Excited Square-Wave Transistor Power Oscillator, G. C. Uchrin, W. O. Taylor. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 43, no. 1, Jan. 1955, p. 99.

4. Variable Frequency Magnetic-Coupled Multivibrator, R. L. Van Allen. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, July 1955, pp. 356–61.

5. Fixed, Composition (Insulated) Resistors *Military Specification MIL-R-11B*, Mar. 9, 1956.

Graphical Evaluation of Magnetic Amplifier Performance Based on Constant-Voltage Reset Test

D. NITZAN
STUDENT MEMBER AIEE

Synopsis. A half-wave voltage-reset magnetic amplifier with a switching transistor in the control circuit is used as a core tester. Families of magnetizing current versus the magnitude of induced voltage with reset flux as a parameter are recorded. A graphical construction similar to the conventional one in the design of electronic amplifiers can yield the core behavior when operating in a 2-core magnetic amplifier, regardless of the magnitude of the control resistance.

THE CHARACTERISTICS of grain-oriented nickel-iron tape cores are described by Feth[1] in terms of the instantaneous values of the net mmf (magnetomotive force) F, the rate of change of flux $\dot{\phi}$, and the flux ϕ. It is assumed that the magnetic state of the resetting core is uniquely determined by these three quantities, provided the following restrictions are fulfilled: the core is initially in saturation, there are no radical changes in F, and there is a monotonic variation of $\dot{\phi}$. The relationship between F, $\dot{\phi}$, and ϕ is then assumed of the form

$$F=F_{avg}(\dot{\phi})+a_1\phi$$

in which F_{avg} is the average mmf corresponding to a full reset from positive to negative saturation. By resetting the core from positive to negative saturation by means of a variable amplitude and variable frequency square-wave supply, the average mmf F_{avg} versus $\dot{\phi}$ is plotted, and this curve is taken as F versus $\dot{\phi}$ at $\phi=0$. The curves of F versus $\dot{\phi}$ for other values of ϕ are determined from the calculated value of a_1.

In this paper a test method is suggested by which the family of F versus $\dot{\phi}$ curves are determined experimentally for any flux level ϕ. This family of curves will be referred to as the "core characteristics." The core characteristics are then employed to predict the operation of a 2-core self-saturating magnetic amplifier with any value of control resistance, including rectifier unblocking.

Constant-Voltage Reset Test

The circuit is shown in Fig. 1. This is a voltage-reset circuit with a synchronous switching transistor in the control circuit. During the positive half-cycle of normal operation the transistor is in the "on" position and the rectifier is blocked. By keeping the resistance R_c to a minimum

---

Paper 59-789, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 19, 1959; made available for printing April 22, 1959.

D. Nitzan, formerly with the University of California, Berkeley, Calif., is now with Stanford Research Institute, Menlo Park, Calif.

This work was performed in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of California.

The author wishes to thank Prof. H. C. Bourne, Jr. for his valuable guidance and helpful comments.

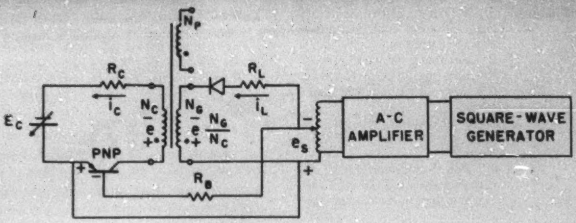


Fig. 1. Constant-voltage reset core test

value, the core is reset from saturation by a constant voltage source. At the end of the positive half-cycle the core reaches a flux level determined by ϕ and the frequency of the square-wave generator. During the next half-cycle the transistor is in the "off" position and the rectifier is conducting. The core is gated by the voltage e_s which is large enough to saturate the core before the end of the negative half-cycle. The value of R_L determines how far the core is driven into saturation. The number of turns N_c, N_G, and N_p may be small enough to enable jig operation.

Typical oscillograms of reset magnetizing current i_c and the induced voltage across the core are shown in Fig. 2. Note that the induced voltage during the negative half-cycle does not have a constant value due to the frequency response of the a-c amplifier. This fact does not invalidate the measurements to be described. The leakage current of the transistor is completely negligible.

The constant value of ϕ can be measured by an average-sensitive voltmeter with very high internal impedance connected across a pickup winding N_p. The corresponding value of F is determined by measuring the instantaneous value of i_c at the end of the reset, positive, half-cycle. This measurement was made by a calibrated oscilloscope, but probably can also be made by a suitable peak-sensitive meter. With a fixed value of E_c, hence constant ϕ, the frequency of the square-wave generator is slowly decreased until the core just reaches negative saturation. Denoting this frequency by f_{min}, resetting the core to any higher flux level ϕ is obtained by increasing the frequency to the value

$$f = f_{min}\frac{2\phi_s}{\phi_s - \phi} \tag{1}$$

This procedure is repeated for other values of ϕ. Evaluation of F versus $\dot\phi$ at $\phi = 0$ is obtained from the conventional measurement of the static hysteresis loop by means of a ballistic galvanometer. The resulting family of curves of a tested core (Magnetic Inc.; 1-mil tape; inside diameter = 1.0 inch; outside diameter = 1.75 inch; height = 0.5 inch; $l_{Fe} = 10.96$ centimeters; $A_c = 0.909$ square centimeter) is shown in Fig. 3. The curve corresponding to $\phi = \phi_s$ is obtained by extrapolating the curves ϕ versus F to $\phi = \phi_s$. The variations in slopes among the curves stem from effects such as macroscopic eddy currents which introduce terms of the nature of $\phi\dot\phi$.

Graphical Construction to Predict Magnetic-Amplifier Performance

Several core testers have been devised in order to obtain the information with which the performance of a 2-core self-saturating magnetic amplifier can be pre-

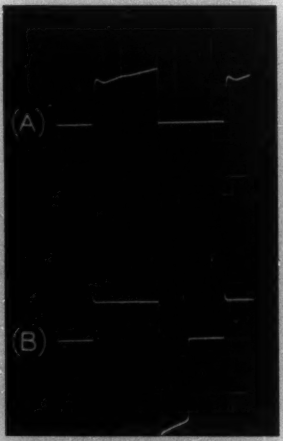


Fig. 2. Oscillograms of reset magnetizing current i_c and induced voltage $N_p\dot\phi$ in a tested core

A—i_c
B—$N_p\dot\phi$

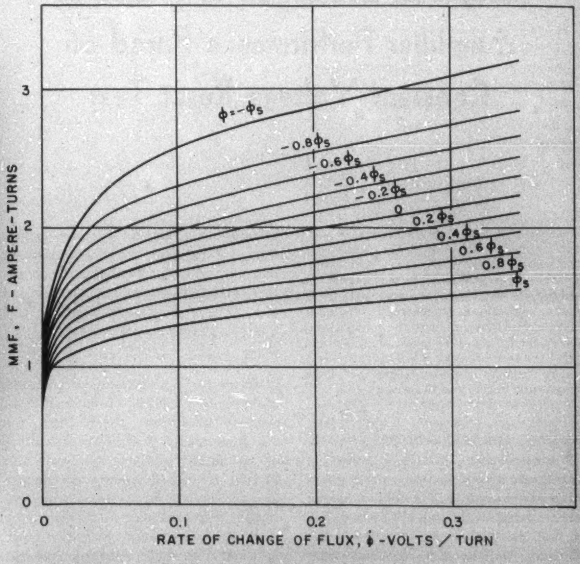


Fig. 3. Core characteristics of a tested core

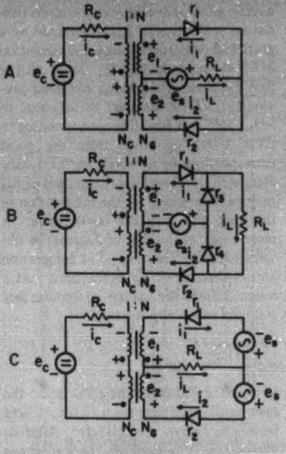


Fig. 4. Three types of self-saturating magnetic amplifier

A—Doubler circuit
B—Bridge circuit
C—Center-tap circuit

dicted.³ Unfortunately, due to the interaction between the two cores, good correlation is limited to extreme conditions of control impedance, e.g., the use of the data from Roberts' constant-current flux-reset core tester³ when R_c is very large.

An attempt is made here to predict the operation of a 2-core self-saturating magnetic amplifier with any value of control resistance by a graphical construction based on the core characteristics. The three types of this circuit are shown in Fig. 4. It is first assumed that the gate rectifier of the resetting core is completely blocked during the whole reset half-cycle. This can be achieved by switched transistors for any value of R_c. Later on this restriction will be removed. Under this condition the common equivalent circuit diagram of the doubler, bridge, and center-tap circuits is shown in Fig. 5. The gate circuit resistance R_o is equal to the sum of the resistances of the load, gate winding, conducting rectifier (*s*), and supply source e_s. Assuming steady-state operation, during mode *I* core 2 gates from $\phi=\phi_r$ at $t=0$ to $\phi=\phi_s$ at $t=\alpha/\omega$ under the voltage

$$e_{2I}=e_s-2i_{mI}R_o \tag{2}$$

and core 1 resets from $\phi=\phi_s$ at $t=0$ to $\phi=\phi_\alpha$ at $t=\alpha/\omega$ under the voltage

$$e_{1I}=e_s-Ne_c-i_{mI}(N^2R_c+2R_o) \tag{3}$$

in which i_{mI} is the instantaneous magnetizing current during mode *I*, referred to the gate circuit and $N=N_G/N_c$. It is assumed in equations 2 and 3 that the difference between the magnetizing currents for the resetting and gating cores produces negligible voltage drop through R_o compared to e_s. During mode *II* the voltage across core 2 is

$$e_{2II}=e_{\text{sat}} \tag{4}$$

and the voltage across the resetting core 1 is

$$e_{1II}=e_{\text{sat}}-Ne_c-i_{mII}N^2R_c \tag{5}$$

in which e_{sat} is the small voltage induced in the saturated core as a result of the variations of its net mmf and i_{mII} is the instantaneous magnetizing current of the resetting core in mode *II*, referred to the gate circuit. The inclusion of e_{sat} is especially important when R_c is small.

For the graphical construction that follows the core characteristics in terms of *F* versus ϕ are referred to the gate circuit in terms of i_m versus *e* by the following conversion factors:

$$i_m=F/N_G \tag{6}$$

and

$$e=N_G\phi \tag{7}$$

An alternative way, which would probably be used in design procedure, is to normalize equations 2 through 5 in terms of volts per turn, mmf, and ohms per turn squared. In this way the conversions indicated by equations 6 and 7 are eliminated and the same plot of core characteristics can be used for any circuit parameters.

SQUARE-WAVE EXCITATION

The case of square-wave excitation and direct-control voltage is discussed first. In this case $e_s=E_s$, $e_c=E_c$ and $e_{\text{sat}}=0$. Equations 2, 3, and 5 become

$$E_s=e_{2I}+2i_{mI}R_o \tag{2A}$$

$$E_s-NE_c=e_{1I}+i_{mI}(N^2R_c+2R_o) \tag{3A}$$

and

$$-NE_c=e_{1II}+i_{mII}N^2R_c \tag{5A}$$

The control voltage $E_{c(0)}$ corresponding to $\alpha=0$, i.e., the minimum value of E_c at which the core remains in saturation is determined from equation 5(A) by letting $e_{1II}=0$. The magnetizing current $i_{mII}=I_{mo}$ is the value of i_m at $e=0$ and $\phi=\phi_s$. At full reset ($\alpha=\pi$) equation 2(A) permits determination of the variation of i_{mI} and e_{2I} versus time by a graphical construction similar to the conventional

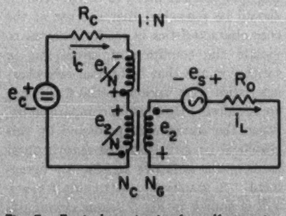


Fig. 5. Equivalent circuit of a self-saturating magnetic amplifier with ideal rectifier blocking

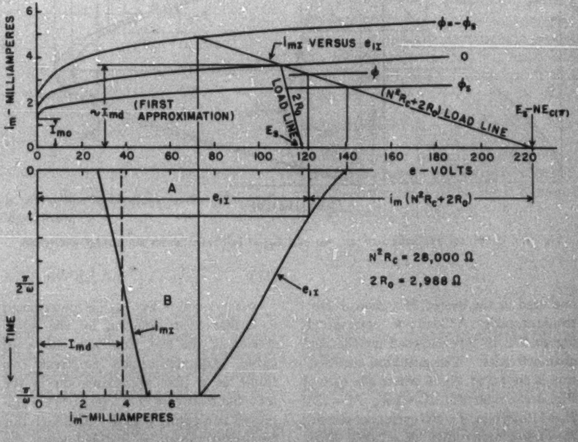


Fig. 6. Graphical evaluation of e_{1I}, i_{mI}, and I_{md} at full reset under square-wave excitation

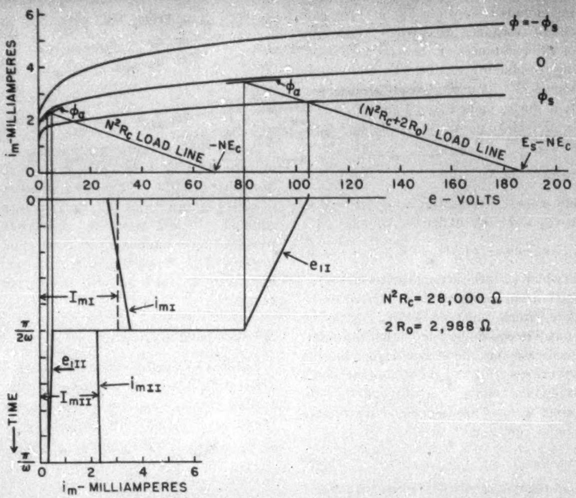


Fig. 7. Graphical evaluation of e_1 and i_m at $\alpha=\pi/2$ under square-wave excitation

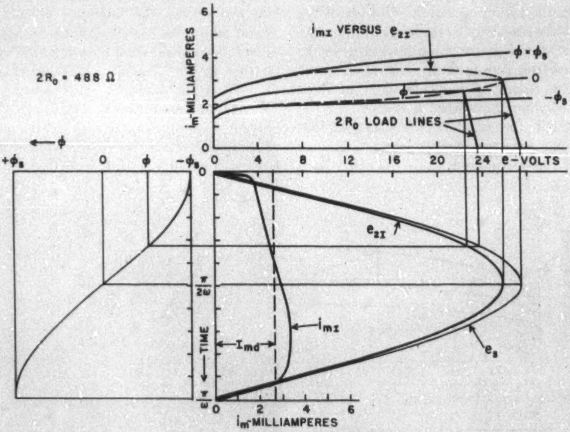


Fig. 8. Graphical evaluation of e_{2I}, i_{mI} and I_{md} at full reset under sinusoidal excitation

one used in the design of a class *A* electronic amplifier: A "load line" corresponding to $2R_o$ is superimposed on the core characteristics. This graphical construction is shown in Fig. 6 where the core of Fig. 3 is employed; $N_G=550$; $N_c=275$; $E_s=119.5$ volts; $f=400$ cycles per second; $R_o=1{,}494$ ohms (Ω); $R_c=7{,}000$ ohms. The average value of i_{mI} from $t=0$ to $t=\pi/\omega$, denoted by I_{md}, is determined as a first approximation by the intersection of the load line with the $\phi=0$ curve. Another reason to treat this result as a first approximation is the fact that reset core characteristics are employed in a gating period. Although this treatment might be justified when the core is not underdesigned, and thus yield a very good approximation, a more consistent evaluation of I_{md} is obtained by a graphical construction based on equation 3(A), as shown in Fig. 6. Based on the first approximation of I_{md} obtained previously, NE_c is assumed to be equal to $-I_{md}N^2R_c$. A load line corresponding to $(N^2R_c+2R_o)$ is drawn from the point corresponding to E_s-NE_c on the *e* axis. Here again a second approximation of I_{md} may be determined by the intersection of the load line with the $\phi=0$ curve. A more exact evaluation of I_{md} can be obtained from a point-by-point integration and evaluation of e_{1I} versus time. At a selected time *t* the voltage e_{1I} and the flux ϕ should satisfy the relation

$$\frac{A}{A+B}=\frac{\phi_s-\phi}{2\phi_s}$$

in which *A* and *B* are the areas under the curve e_{1I} versus time from $t=0$ to *t*, and from *t* to $t=\pi/\omega$ respectively. After e_{1I} versus time is determined, the waveshape of i_{mI} from $t=0$ to $t=\pi/\omega$ is easily plotted directly from the core characteristics, as shown in Fig. 6. The average value of this waveshape is the final value of I_{md} and its product with $-NR_c$ yields the control voltage $E_{c(\pi)}$ corresponding to $\alpha=\pi$.

If the firing angle α is between 0 and π, a cut-and-try graphical construction is employed. The amount of work involved depends on the degree of accuracy sought. As a first approximation the corresponding value of $-NE_c$ is

$$-NE_c\cong N^2R_c\left(I_{md}\frac{\alpha}{\pi}+I_{mo}\frac{\pi-\alpha}{\pi}\right)\qquad(8)$$

The graphical construction for mode *I* is similar to the previous one. The reset flux ϕ_a at the end of mode *I* is selected to be slightly above the reset flux ϕ_r at the end of mode *II*. The latter is calculated from the expression

$$\phi_r=\phi_s\left(1-2\frac{\alpha}{\pi}\right)\qquad(9)$$

The graphical construction during mode *II* is based on equation 5(A) with a load line corresponding to N^2R_c. Here it should be remarked that the use of the core characteristics in mode *II* seems to violate the condition of initial flux in saturation and that additional core characteristics with various initial flux levels are required. However, a comparison between the waveshapes predicted by this graphical construction and experimental oscillograms indicates satisfactory agreement. A graphical construction for $\alpha=\pi/2$ is shown in Fig. 7 (same circuit parameters as for Fig. 6). Let mode-averaged magnetizing currents be defined as

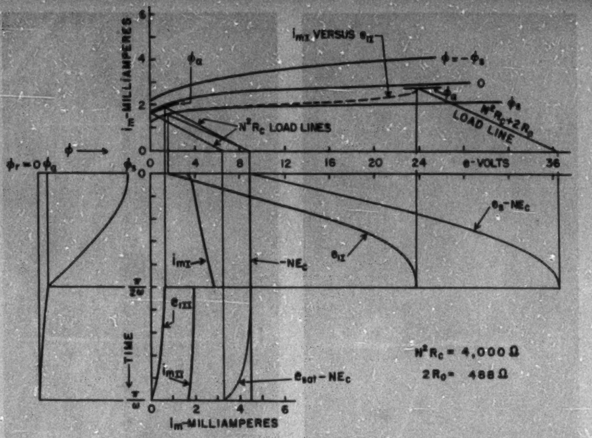


Fig. 9. Graphical evaluation of e_1 and i_m at $\alpha=\pi/2$ under sinusoidal excitation

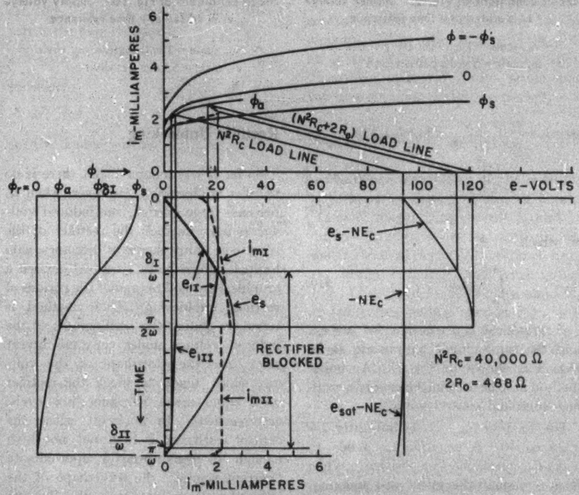


Fig. 10. Graphical determination of the periods of rectifier unblocking under sinusoidal excitation in a bridge circuit

$$I_{mI}=\frac{1}{\alpha}\int_0^\alpha i_{mI}d\omega t \qquad (10)$$

and

$$I_{mII}=\frac{1}{\pi-\alpha}\int_\alpha^\pi i_{mII}d\omega t \qquad (11)$$

The resulting values of I_{mI} and I_{mII} should agree with the assumed value of $-NE_c$ according to

$$-NE_c=N^2R_c\left(I_{mI}\frac{\alpha}{\pi}+I_{mII}\frac{\pi-\alpha}{\pi}\right) \qquad (12)$$

In addition, the ratio between the area under e_{III} to the area under e_{II} should equal to $(\phi_a-\phi_r)/(\phi_s-\phi_a)$. If necessary, a correction to the assumed $-NE_c$ and ϕ_a is made and a similar graphical contruction is repeated until these conditions are satisfied. Practically it was found that two trials were adequate.

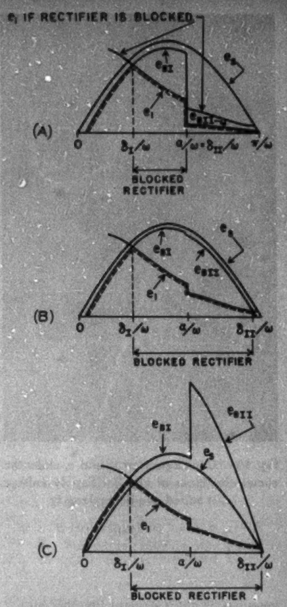


Fig. 11. Rectifier unblocking in self-saturating magnetic amplifiers under sinusoidal excitation

A—Doubler circuit
B—Bridge circuit
C—Center-tap circuit

Sinusoidal Excitation

The case where e_s is sinusoidal is more involved because both e_s and e_{sat} vary with time. A graphical construction to predict the waveshape of i_{mI} at full reset, and hence I_{md}, is based on equation 2 and is shown in Fig. 8 (core of Fig. 3; $N_G=550$; $N_C=275$; $e_s=\frac{\pi}{2}$ 17.5 sin ωt volts; $f=60$ cycles per second; $R_o=244$ ohms; $R_c=1{,}000$ ohms; d-c control voltage). The variations of ϕ with time is obtained by integrating equation 2

$$\phi=-\phi_s+\frac{1}{\omega N_G}\left[\frac{\pi}{2}E_s(1-\cos\omega t)-2R_o\int_0^{\omega t} i_{mI}d\omega t\right] \qquad (13)$$

The total flux change completed by $t=\pi/\omega$ is

$$\Delta\phi_r=2\phi_s=\frac{1}{\omega N_G}\pi(E_s-2I_{md}R_o) \qquad (14)$$

Note that in contrast to the case of square-wave excitation the locus of i_{mI}

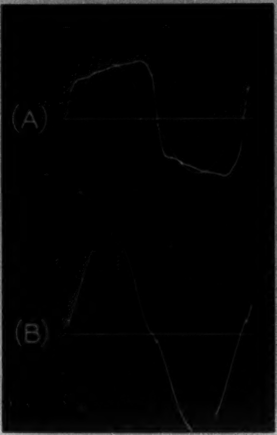


Fig. 12. Oscillograms of i_m and e_1 under the circuit conditions of Fig. 8. Supply voltage e_s is added as a time reference

A—i_m, scale = 1 milliampere per division
B—e_1, scale = 5 volts per division

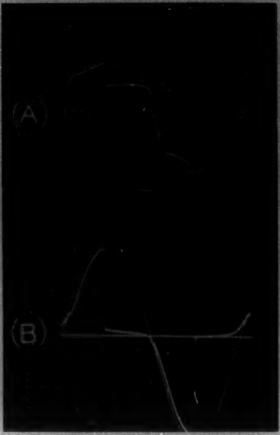


Fig. 13. Oscillograms of i_m and e_1 under the circuit conditions of Fig. 9. Supply voltage e_s is added as a time reference

A—i_m, scale = 1 milliampere per division
B—e_1, scale = 5 volts per division

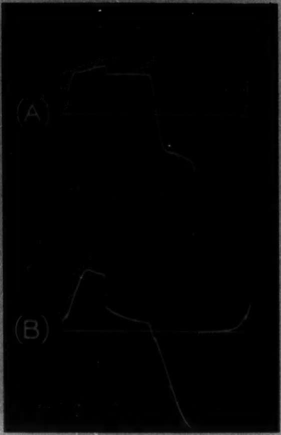


Fig. 14. Oscillograms of i_m and e_1 unde circuit conditions of Fig. 10. Supply voltage e_s is added as a time reference

A—i_m, scale = 1 milliampere per division
B—e_1, scale = 5 volts per division

Vol.
Pt. 1
195

versus e_{2I} is not a straight line but curved as shown in Fig. 8.

For a firing angle $o<\alpha<\pi$ a first estimate of $-NE_c$ is computed from equation 8 and the graphical construction for modes I and II is based on equations 3 and 5 respectively. The waveshape of e_{sat}, equation 5, is independent of R_c and can be obtained from the "butterfly loop" of the sinusoidal-current core tester.[2] Integration of equation 3 yields ϕ versus time during mode I

$$\phi_I=\phi_s-\frac{1}{\omega N_G}\left[\frac{\pi}{2}E_s(1-\cos\omega t)-NE_c\omega t-\right.$$
$$\left.(N^2R_c+2R_o)\int_0^{\omega t}i_{mI}d\omega t\right] \quad (15)$$

The flux $\phi_I=\phi_\alpha$ at $t=\alpha/\omega$ is

$$\phi_\alpha=\phi_s-$$
$$\frac{\frac{\pi}{2}E_s(1-\cos\alpha)-[NE_c+I_{mI}(N^2R_c+2R_o)]\,\alpha}{\omega N_G} \quad (16)$$

Similarly ϕ versus time during mode II is obtained by integrating equation 5

$$\phi_{II}=\phi_\alpha-\frac{1}{\omega N_G}\left[\int_\alpha^{\omega t}e_{sat}d\omega t-\right.$$
$$\left.NE_c(\omega t-\alpha)-N^2R_c\int_\alpha^{\omega t}i_{mII}d\omega t\right] \quad (17)$$

From equations 2 and 17 $\phi_{II}=\phi_r$ at $t=\pi/\omega$ is

$$\phi_r=\phi_\alpha-\frac{\pi-\alpha}{\omega N_G}(E_{sat}-NE_c-I_{mII}N^2R_c)$$
$$=\phi_s-\frac{1}{\omega N_G}\left[\frac{\pi}{2}E_s(1-\cos\alpha)-2I_{mI}R_o\alpha\right] \quad (18)$$

in which

$$E_{sat}=\frac{1}{\pi-\alpha}\int_\alpha^\pi e_{sat}d\omega t \quad (19)$$

A graphical construction for $\alpha=\pi/2$ and the same circuit parameters as for Fig. 8 is shown in Fig. 9. A similar method may be used in the case of a rectified-sinusoidal control voltage.

Up to $t=\pi/2\omega$ $e_{sat}>0$, but after $t=\pi/2\omega$ $e_{sat}<0$. A negative e_{sat} tends to make e_{1II} equation 5, less positive. When R_c is very small this effect will make e_{1II} negative during the end portion of the cycle. During this period the core has already started to gate along the minor loop corresponding to the reset flux. This amounts to an extension of the core characteristics into the negative region of ϕ, starting with positive F and ending with negative F. The voltage e_{sat} has an appreciable effect on decreasing $|E_c|$ but has a negligible effect on the total flux reset, in other words, a shift to the right of the transfer characteristic. This phenomenon reaches the point where it has maximum effect when the firing angle is equal to $\pi/2$.

Rectifier Unblocking

So far it was assumed that there is no rectifier unblocking. However, when R_c becomes large enough, the induced voltage e_1 may unblock the rectifier of the reset core, and the reset becomes gate limited. Finzi and Critchlow[4] devised a graphical method to predict the periods of rectifier unblocking. Their method is based on sliding the oscillograms of the induced voltage under constant-current reset, over the waveshape of the voltage which tends to block the rectifier until equal areas, i.e. same flux levels, are reached. In a circuit where the control resistance is high but not high enough to constitute an approximate "current source," the waveshape of the induced voltage may look like the one under constant-current reset but is certainly not identical with it. Under this circumstance the afore mentioned method cannot be applied, unless proper oscillograms of the induced voltage have been supplied.

The use of the graphical construction in conjunction with the core characteristics yields the waveshape of the induced voltage regardless of R_c. This method, therefore, can be used to predict the periods of rectifier unblocking with any value of control resistance.

Referring to the gate circuits of the re-

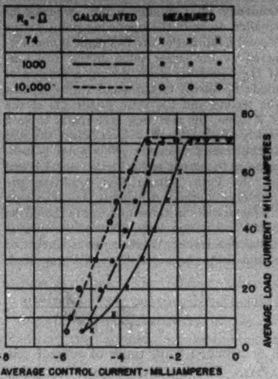


Fig. 15. Transfer characteristics of a bridge circuit

set core in Fig. 4, the blocking voltage across the gate rectifier is

$$e_r = e_B - e_1 \qquad (20)$$

in which

$$e_B = e_s - i_L R_L \qquad (21A)$$

for the doubler circuit

$$e_B = e_s - e_f \qquad (21B)$$

for the bridge circuit and

$$e_B = e_s + i_L R_L \qquad (21C)$$

for the center-tap circuit.

The symbol e_f stands for the forward voltage drop across the conducting complementary rectifier. The saturating rectifier stays blocked as long as $e_r \geqslant 0$. (The equivalent circuit for the forward characteristic of a silicon rectifier may be represented by a resistance R_f in series with a constant voltage E_u. In such a case the rectifier stays blocked as long as $e_r \geqslant -E_u$.) This period can be found graphically by subtracting e_1 from e_B. An example is shown in Fig. 10 for the bridge circuit with identical circuit parameters as in Fig. 9, except that R_c is ten times larger. The waveshape of e_{1I} versus time is plotted "backwards" from $t=\alpha/\pi$ towards $t=0$ until at $t=\delta I/\omega$, where $0<\delta I/\omega<\alpha/\omega$, $e_{1I}=e_B$. Similarly the waveshape of e_{1II} is plotted "forwards" from $t=\alpha/\omega$ towards $t=\pi/\omega$ until at $t=\delta II/\omega$, where $\alpha/\omega<\delta II/\omega<\pi/\omega$, $e_{1II}=e_B$. During the period $\delta I/\omega \leqslant t \leqslant \delta II/\omega$ the rectifier stays blocked. It is obvious from equations 21 that the worst case of rectifier unblocking is in the doubler circuit, especially in mode *II*, whereas the center-tap circuit has the least rectifier unblocking, shown schematically in Fig. 11.

The expression for i_1 and e_1 during the periods of rectifier unblocking must be determined in conjunction with the rest of the circuit. During these periods the control current $|i_c|$ must increase by the amount $|Ni_1|$ in order to satisfy the constraints of net mmf of the resetting core. This effect, as is well known, would shift the transfer characteristic to the left.

Experimental Verification

Checking the validity of the resulting waveshapes of Figs. 8, 9, and 10, oscillograms of magnetizing current i_m and induced voltage e_1 were photographed under identical circuitry conditions, using the bridge circuit. These oscillograms, Figs. 12, 13, and 14, show close similarity to the corresponding ones in Figs. 8, 9, and 10, respectively. Note that there is slight phase shift in i_m and a slight delay time in the switching of the complementary rectifiers caused by e_{sat}.

Repeating the graphical construction for other values of α and R_c results in calculated transfer characteristics which are compared with experimental ones in Fig. 15.

Conclusions

The inherent nature of a ferromagnetic core is expressed by its core characteristics which are obtained from a constant-voltage reset test. The core characteristics enable one to predict the variations of F, $\dot{\phi}$, and ϕ with time for any value of R_c by means of a graphical method. The resulting information helps to estimate the performance of the designed magnetic amplifier, including rectifier unblocking. This method has been demonstrated for the 2-core self-saturating magnetic amplifier with square-wave and sinusoidal excitation. The amount of the cut-and-try work involved depends on the required accuracy. If the cores are underdesigned the same method can be applied, except for a slight modification: At $t=\pi/\omega$ the full reset flux ϕ_π is higher than $-\phi_s$, and hence $(\phi_s-\phi_\pi)$ should supersede $2\phi_s$ wherever full reset is considered.

References

1. Core-Reset Functions in Magnetic Amplifier Analysis, G. C. Feth. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, Sept. 1958, pp. 503–09.

2. Present Methods of Core Evaluation, Grading and Matching, R. W. Roberts. *Proceedings*, "Special Technical Conference on Magnetic Amplifiers," *AIEE Special Publication T-86*, July 1955.

3. Magnetic Characteristics Pertinent to the Operation of Cores in Self-Saturating Magnetic Amplifiers, R. W. Roberts. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 73, 1954 (Jan. 1955 section), pp. 682–90.

4. Dynamic Core and Magnetic Amplifier Performance, L. A. Finzi, D. L. Critchlow. *Ibid.*, vol. 76, May 1957, pp. 229–40.

◆

Discussion

G. C. Feth (General Electric Company, Schenectady, N. Y.): Use of the "constant-voltage reset test" described in this paper to obtain information on the instantaneous values of mmf, flux, and time-rate of change of flux in a core, resetting from saturation, should yield somewhat more complete and accurate information about the core characteristics than that contained in the parameters of the core-reset functions which were described in reference 1 of the paper. In evaluating such parameters, some accuracy is sacrificed in condensing the data so that an analytical formulation is possible; but the resulting analytical formulation is then readily utilized in a digital computer for speedy solution of problems, avoiding the more time-consuming graphical constructions. It should be emphasized, however, that regardless which of the two methods of obtaining data on the core characteristics is used there is still difficulty in calculating the control characteristic of the 2-core full-wave self-saturating amplifiers, center-tap, doubler, bridge, with sinusoidal excitation voltage or with square-wave voltage and capacitive load except when the amplifiers are controlled essentially from a current source. This is true because after the load current in the saturated core has reached its crest and decreases toward zero a sizeable voltage is induced in the control winding in a direction to prevent further resetting of the unsaturated core. That core then experiences a varying mmf which usually, except for nearly current-source control, ranges from the resetting branch of the hysteresis loop, through the static loop to the ascending branch of some minor loop.

Any information which may be obtained about the core from the constant voltage reset test does not apply here; hence any meaningful analysis in terms of that information is precluded. Moreover, although the amounts of flux change involved may be but a few per cent of ϕ_s, the control current may range from its maximum absolute value to a value of opposite sign with magnitude perhaps half its maximum. This control current excursion may last as long as one half of the half-cycle, and with no means of predicting the exact waveform, no accurate appraisal of the average control current can be made.

D. Nitzan: The author wishes to thank Dr. Feth for his interesting comments about some points which, it seems, are not emphasized enough in the paper. In evaluating the inaccuracy found in the analytical representation of the core characteristics in reference 1 compared with the core characteristics used in this paper, two factors should be distinguished. The first one stems from the assumed core model according to which F varies linearly with ϕ. The second factor depends on how closely the measured F_{avg} is approximated by the chosen functional form. A solution using a digital computer is certainly faster than by the graphical construction. However, which one is preferable depends upon the operational costs and upon the availability of the computer and the degree of accuracy required. An alternate way to insert the core characteristics into a computer program is by tabulation of the measured curves.

The effect of the load current in the saturated core on the resetting core is taken into account by adding the voltage source e_{sat}, which is independent of R_c, to $-NE_c$. If the operating point on the load line in mode *II* is within the first quadrant of the F-ϕ plane, core 1 would continue to reset and the information obtained from the constant voltage reset test of Fig. 1 is certainly useful. To fulfill this condition the control circuit does not necessarily have to constitute a current source. A wide range of finite values of R_c/N_c^2, especially when α is large, is under this category (see, for example, Figs. 9 and 13). Note that the tendency of a smaller $-NE_c$, when R_c is smaller, to shift the operating point to the second quadrant of the F-ϕ plane is counteracted by the larger slope of the load line. If the operating point during a portion of mode *II* moves to the second quadrant or the second and the third quadrants of the F-ϕ plane, the resetting core starts to gate slowly along the minor loop before the end of the reset half-cycle. This point is indicated in the paper and is further emphasized by Dr. Feth. Under this condition the graphical method still holds, but the core characteristics in these quadrants should be based on minor loops test with various initial flux levels. Further investigation to obtain these characteristics is still required.

Correlation of Measured and Calculated Substation Ground Grid Resistance

A. L. KINYON
MEMBER AIEE

IN THE past, substation ground grids have been designed by rule-of-thumb methods. There is reason to believe that as a result of this, many were greatly overdesigned whereas others gave a sense of security that was questionable.

In recent years methods have been developed which permit a more scientific approach to grid design. Using these methods, and with adequate knowledge of soil resistivity, it is possible to predict, at least approximately, the resistance to earth of the total grid before actual construction.

Several methods have been developed by various authors[1–6] for computing grid resistance for a given resistivity and grid geometry. Although some approximations are involved in relating most practical grids to the geometrical forms assumed by these methods, the results will generally be sufficiently accurate for practical purposes.

However, the resistivity value utilized in these formulas is often the source of errors of considerable magnitude since the formulas are valid only for soils of uniform resistivity, a condition not ordinarily encountered in most soils to any appreciable depth. Consequently, the resistivity value established at a given site by the survey becomes an important factor in obtaining accuracy in the proposed design.

Resistivity values measured in the field, using conventional instruments, represent an average or apparent resistivity of a geometric solid of earth, the dimensions of which are dependent upon the test conditions chosen. Such a solid may contain several individual layers or portions, each of which may have a resistivity vastly different than that of other parts.

Reasonably successful work has been done in the calculation of resistance using a single resistivity value which, in turn, is determined from those of several soil layers. The methods require extensive field measurements and laborious calculations and are therefore impractical except for unusual cases.

In general the complications involved make this a problem for which it would be highly desirable to find a reasonably accurate solution based upon the direct use of the apparent resitivity as measured in the field. It appeared that some progress may be made toward such a solution by approaching it from the study of various trends observed within the field test data taken over a period of years.

Method of Test

Resistivity Measurements

The resistivity measurements were made by the "four point" method using a conventional ground resistance instrument. Briefly, this consists of driving four probes into the earth in a straight line, at equal spacings, circulating a current i between the outer two probes C, and measuring a potential v across the inner two probes P (see Fig. 1).

The instrument is so scaled and calibrated that the indication is read in terms of v/i or the mutual resistance in ohms directly. The resistivity ρ is then calculated from the equation (modified form for relatively short probes)

$$\rho = 2\pi AR \qquad (1)$$

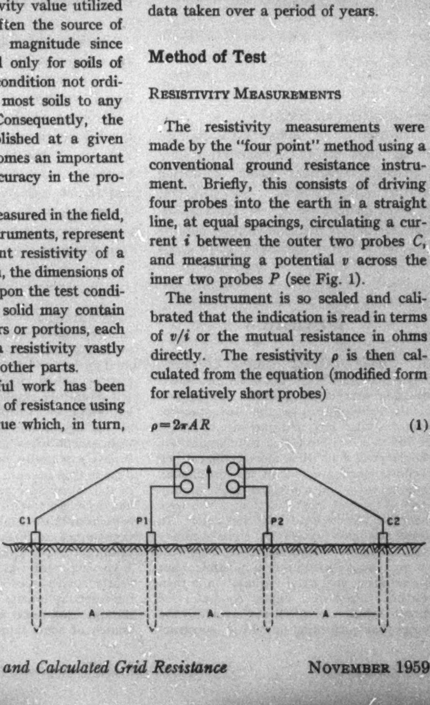


Fig. 1. Method for testing soil resistivity

Paper 59-805, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing April 27, 1959.

A. L. Kinyon is with the Bonneville Power Administration, Vancouver, Wash.

The author expresses sincere thanks and appreciation for the critical review and suggestions given by members of the Bonneville Power Administration Staff and by members of Task Force on Section 5.00 and 6.00 of the 1958 Subcommittee on Ground Resistance and Potential Gradient Measurements of the AIEE Special Instruments and Auxiliary Apparatus Committee. The material for this report is based upon a summary of the ground resistance and resistivity data accumulated during the past 18 years on the Bonneville Power Administration System.

Vol.
Pt. 1
1959

where

A = distance between the probes in cm (centimeters)
R = reading of the instrument in ohms
ρ = resistivity in ohms per cm cube

It has generally been assumed that, inasmuch as the grid and the ground rods were usually buried within the first 10 feet of depth, measurements encompassing this depth at the proposed substation site should be reliable for use in calculating resistance. Resistivity measurements were, therefore, made at the proposed site using 5-, 10-, and 20-foot probe spacings. The reading taken at any given spacing was considered to represent the average resistivity of the soil between the surface and a depth equal to the spacing. These data were then used for design purposes.

Resistance Measurements

Ground resistance of the installed ground grids were measured with a ground megohmeter. Sufficient data were taken to plot a complete curve of mutual resistance versus potential probe distance from the substation. This provided a graphic record for reference in future tests. These tests were performed immediately after construction and again approximately one year later after soil settling and seasonal changes had occurred. Subsequent tests were repeated at, approximately, 5-year intervals.

Information in the records which appeared to be of value was as follows:

1. Review of the data taken during the earlier years showed that measurements made with three different makes of instruments were in close agreement when measuring resistance but quite often in wide variance when measuring resistivity. Limited studies made at that time indicated that the errors were due to potential probe resistances. Some of the instruments used were null-type and were free from these effects but many of the measurements had been made with both the null and the "moving-coil ohmeter types."
2. Resistance of substations decreased with time because of weather effects and soil settling. This decrease may amount to 50% or more of the original value and occurred largely during the first year or two after construction. The resistance did not appear to vary greatly with seasonal changes after the first year or two.
3. Soil resistivity values obtained at 5-, 10-, and 20-foot spacings may change radically (500–1,000% or more) with seasonal weather changes. Usually these changes were slightly less pronounced at 20-foot spacings than at the 10- or 5-foot spacings.

As a result of the foregoing, additional data were obtained by taking the following steps:

1. Three types of instruments were tested to determine the effects of high probe resistances.
2. To eliminate the difference in measured and calculated resistance caused by a possible difference in moisture content when the resistivity survey was made and when the resistance measurement was made, resistivity measurements were again made in the proximity of substations during one season at the same time the periodic resistance measurements were conducted.
3. Using the previously mentioned data, an attempt was made to correlate measured resistance with calculated resistance values. Attention was also given the possibility of determining the relationship of the measured resistance to each of the resistivity values measured at the 5-, 10-, and 20-foot spacings.

Effect of Probe Resistance

Comparative tests were made on two null-type and one moving-coil ohmeter-type instruments by inserting resistance in series with the probes during resitivity measurements, under field conditions.

Resistance connected in series with the current probe circuit affected only the sensitivity of all instruments as presumed. Resistance connected in series with the potential probes caused no appreciable difference in the indication of null-type but caused a sharp decrease in the indication of the moving-coil ohmeter type (see Fig. 2).

Joint Resistivity and Resistance Measurements

Tests made at 47 substations during the 1958 season included resistance of the substation grid and resistivity taken the same day at one or more positions within the proximity of the grid. The resistance was measured with a ground megohmeter using the conventional method and plotting a mutual resistance curve from which the resistance was determined. Resistivity measurements included readings at 5-, 10-, and 20-foot probe spacings using the null-type instrument.

Comparison of Calculated and Measured Resistance

Several of the formulas developed for calculation of resistance from resistivity values were considered. The choice of formula appeared to be of much less importance than the value of resistivity used. The formula developed by Laurent was used primarily because of its applicability and its simplicity. This formula is

$$R=\frac{\rho}{4r}+\frac{\rho}{L} \tag{2}$$

where

R = resistance of the substation grid
ρ = resistivity of the soil in the proximity of the substation in ohms per cm
L = total length of buried conductor in cm
r = radius of a circle having an area equal to that of the substation grid

This equation is based upon the assumption that the grid is located at or very near the surface of the ground. The grids measured generally consisted of a buried conductor around the perimeter with sufficient cross-mesh to ground all equipment. They were buried approximately 1 foot to 18 inches deep with a scattering of ground rods driven 6 to 10 feet deep.

The resistivity data assembled from tests at the 47 substations showed wide variations between the values at 5-, 10-, and 20-foot spacings. Ordinarily, the resistivity decreased in value with increased probe spacing. However, at a few stations the value remained fairly constant and at others it increased with increased spacing. Resistance calculations were made using an average of the 5-, 10-, and 20-foot resistivity values. Comparison of the calculated and measured resistance showed that about one half of the calculated values fell within ±300% of those measured, with the remaining results even more widely scattered.

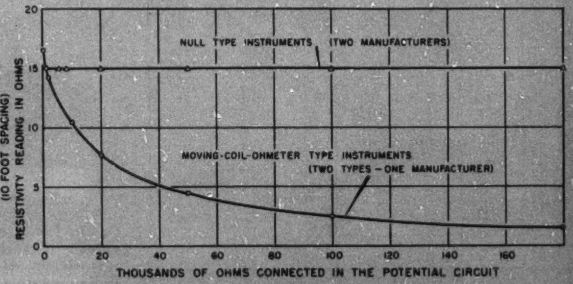


Fig. 2. Effect of potential probe resistance

Table I. Comparison of the Measured and the Calculated Ground Resistance

Number of Substations	(1)	(11)	(13)	(17)	(35)	(44)	(47)	(51)	(59)	(60)	(70)
Equivalent Diameter (Feet)	360	99	639	355	134	250	677	304	391	79	119
Measured Resistance (Ohms)	0.04	1.30	0.10	1.20	1.63	0.12	0.33	0.12	7.30	0.44	0.82
Probe Spacing (Feet)	**Calculated Resistance Based on Resistivity for Various Probe Spacings**										
5	0.21	11.25	0.68	4.64	16.97	4.79	7.96	0.18	89.77	24.69	4.15
10	0.16	7.14	0.61	6.34	5.90	2.03	12.82	0.09	111.16	8.24	1.36
20	0.12	5.41	0.79	3.12	2.48	0.85	19.04	0.03	93.36	3.04	1.00
30	0.07	5.84	0.86	4.01	2.53	0.43	17.13	0.04	78.00	1.71	1.06
40	0.03	3.90	0.77	4.45	2.11	0.23	15.21	*	55.19	1.15	1.07
50	*	3.87	0.70		1.58	0.16	10.99		24.23	*	
60		4.19	0.64	5.26	1.55	0.16	13.13		20.82		1.00
70		4.41			1.33*	0.14	15.35		14.24		
80		2.80	0.45	5.87	1.05*	0.16	10.53		12.68		1.14
90		3.15			0.86*	0.16	11.86		10.09		
100	0.27	3.51	0.64	5.99	0.58*	0.15	16.01		4.29		0.93
150				3.99			8.33		8.39		
200				3.11			9.74		26.89		

* Instrument reading too low to be accurate.

Closer study showed that of the 47 stations only about one fourth could be considered relatively isolated and free from effects of overhead grounds, grounded neutrals, other adjacent grounded substations, city water systems, etc. Data from these isolated stations showed a closer than average relationship between the measured and the calculated resistance. The resistance values calculated from the resistivity readings taken at the 20-foot spacings at these stations were in closer agreement with the measured resistance than those taken at the 10- or 5-foot spacings.

These findings and other similar evidence indicated the existence of an average resistivity value between the surface and some depth greater than 20 feet which, when used in equation 2, would result in a resistance value more closely approximating the measured value than heretofore obtained.

Within the time available at the close of the testing season, additional resistivity measurements were made at 11 of the substations. These measurements were conducted at the 5-, 10-, and 20-foot spacings, then the spacings were progressively increased to a maximum of 200 feet where possible. Table I shows the measured resistance and the resistance as calculated from resistivity values at the various spacings for these 11 substations. The table shows that for most of the substations the calculated resistance values are extremely high in relation to measured values when using resistivity values measured at 5-, 10-, or even 20-foot spacings. Results at about one half the substations show that when resistivity values taken at 50- to 150-foot spacings are used, the calculated resistance closely approximates the measured resistance value. Results at the remaining substations showed that resistance calculated from resistivity at these wider spacings were usually as close or closer in agreement with the measured value than those which were obtained from readings at 20 feet or less.

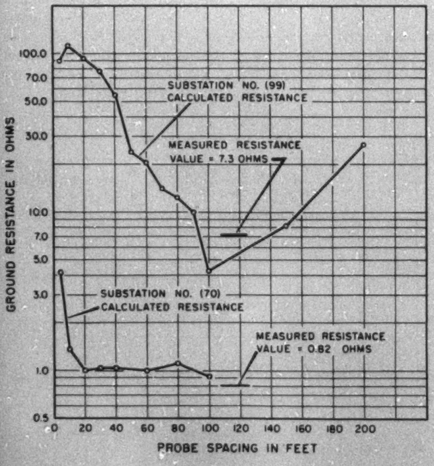


Fig. 3. Calculated ground resistance using resistivity values measured at various probe spacings

These 11 stations were chosen for their geographical location rather than other considerations. Several sources of error are present, examples of which are given in the following.

A minor source of uncertainty and possible error in the data resulted from the necessity of making resistivity measurements outside the substation enclosure rather than within the space actually occupied by the grid. Using the instruments described, resistivity values measured prior to placement of the grid cannot be duplicated after placement. A limited number of resistivity measurements made very near or immediately above buried grids have resulted in values which were considerably below those measured before the grid was buried.

Other sources of error which are present near energized substations are the effects of 60-cycle ground currents and their harmonics, which may enter into the instrument circuitry. Limited investigations indicate that such currents affect the accuracy of the instrument by the additional burden they impose rather than by combining with the odd frequency current supplied by the instrument.

An examination of equation 1 shows that with a given soil resistivity, spacing A can be increased to a value where reading R would be below the range of the instruments. These effects were in evidence at some of the stations even at relatively close probe spacings.

In equation 2, if the radius r is not accurate, the result will vary accordingly. At some of the substations, the overhead grounds are connected, city water lines are either connected or in close proximity, and grounded neutrals are carried from the station. All of these items tend to increase the effective area of the station and, consequently, the calculated radius r. Only a few of these 11 stations could be considered isolated and free from these interfering conditions.

Fig. 3 shows the calculated resistance plotted as a function of the probe spacing for two substations which were known to be relatively free from interference and at which the resistivity readings were

Vol.
Pt. 1
195

sufficiently high to permit relatively wide probe spacings.

Conclusions

The study indicates that improvements can be obtained in the accuracy of the calculated value of grounding grid resistance when utilizing the apparent resistivity as measured with conventional ground resistance instruments. The improvements depend upon a fairly comprehensive resistivity survey where soil resistivity varies widely at different depths near the surface.

It appears that the survey for a proposed substation grid should include a series of tests at two or more locations on the area, depending upon the proposed size. The investigations indicate that resistivity values tend to stabilize at spacings of 50 feet or more. Each series of tests, therefore, should include measurements at probe spacings beginning at 5 or 10 feet and increasing in increments of 10 to 20 feet until the spacings have reached about 100 feet for small to medium size substations, and 200 feet or more for large substations. It is suspected that the maximum spacing required to obtain stable resistivity values may approach the diagonal distance of the proposed grid, instrumentation permitting.

The data indicate that a stable value should be recognized as one which has become relatively constant for several consecutive tests at increased spacings or one which has reached an optimum low value at some spacing beyond which the value may tend to increase again.

It is apparent that refinements in determining the significance of intervening measured values and the maximum spacings and the recognition of the stable value will depend upon experience to be gained by field measurements. This may also depend upon progress in the improvement of instrumentation such as the provision for increased energy into the probe circuit and expansion of the lower range, or both.

The findings discussed in this report indicate progress and give rise to new questions. It is hoped that subsequent testing performed in this field in the future, both at installed grounding grids and at site prior to construction, will contribute to a greater understanding and a more accurate solution of the problem.

References

1. Voltage Gradients Through the Ground Under Fault Conditions, **AIEE Committee Report.** *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 77, Oct. 1958, pp. 669–92.

2. Grounding Principles and Practice, I—Fundamental Considerations on Ground Currents, **R. Rudenberg.** *Electrical Engineering*, vol. 64, no. 1, Jan. 1945, pp. 1–13.

3. Manual on Ground Testing. **James G. Biddle Company**, Philadelphia, Pa., reprint no. 25J, Mar. 1952, pp. 31–34.

4. Earth Conduction Effects in Transmission Systems (book), **E. D. Sunde.** D. Van Nostrand Company, Inc., Princeton, N. J., 1949.

5. Underground Corrosion, *Circular C450*, National Bureau of Standards, Washington, D. C., Nov. 27, 1945.

6. General Fundamentals of Electrical Grounding Techniques (Les Bases Generales de la Technique des Mises a la Terre Dans les Installations Electriques), **P. G. Laurent.** *Bulletin*, Société Francaise des Électricians, Malakoff, France, July 1951.

7. Efficiency of Grounding Grids with Nonuniform Soil, **J. Zaborszky.** *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 74, Dec. 1955, pp. 1230–33.

8. Some of the Fundamental Aspects of Ground Resistance Measurements, **E. B. Curdts.** *Ibid.*, pt. I (*Communication and Electronics*), vol. 77, Nov. 1958, pp. 760–67.

◆

Discussion

R. F. Stevens (Bonneville Power Administration, Portland, Oreg.): Mr. Kinyon's paper sheds valuable new light on a problem that has received intensive recent study by two AIEE working groups. The problem is that of voltage grandients along the ground under fault conditions.

The first of these groups presented the report listed by Mr. Kinyon as reference 1 in February 1958. Subsequently, this material was revised and expanded into what has been called "A Guide to Safety in A.C. Substation Ground," which is now under consideration by the Committee on Standards for possible publication as a separate document. This guide approaches the problem primarily from the point of view of the designer.

Meanwhile, an AIEE Ground Resistance and Potential Subcommittee, of which John Parker is chairman and Mr. Kinyon is a member, is approaching the problem of both ground resistance and gradients from an instrumentation and measurement point of view, and will present its report a little later.

Now it is generally felt that a physician who has confidence in his methods should be willing to have them tried on his family. The same should doubtless apply to a utility engineer, like myself, who has worked on an AIEE project. We therefore started to check the safety of our Bonneville Power Administration ground grids by the principles established by AIEE Working Group 56.1. However, we met a stumbling block almost immediately when we tried to use the measured values of resistance and soil resistivity.

It is well known that the absolute potential to which a ground mat will rise can be computed as a simple product of short-circuit current times station ground resistance. The safety or danger at a station, however, is not usually a matter of this total potential rise, but of the magnitude of local potential differences which can be bridged by a man in or near the station. These local potential differences have no simple direct relation to the station ground resistance. They can conceivably be dangerously high where the resistance is relatively low, or within safe limits where it is very high. They are determined solely by ground current, grid geometry, and resistivity. Thus it is important to have meaningful figures on the effective soil resistivity at a station.

In the case of each of our substations, we had available one reading each for 5-, 10-, and 20-foot probe spacing. We also had ground resistance measurements taken by the fall of potential method, often on a different day from that on which resistivity was measured.

In some cases, the readings were reasonably consistent for the three probe spacings, showing the ground to be fairly homogeneous down to a depth of about 20 feet. In other cases, they varied considerably, and we faced the question as to which value to use in our computations.

As previously noted, the station resistivity does not enter into the computation of the local potential differences. However, by using the measured resistivity and known grid geometry, we were able to compute a value of station grid resistance by any one of various methods described in the literature and compare this with the resistivity as measured. Where agreement is good, this provides a partial check on the validity of both measurements. Where agreement is very poor, one measurement or the other must be invalid.

Actually, agreement was very poor in many cases, with calculated values often not merely differing by a few per cent, but being perhaps four or five times the measured values. Application of the methods of calculations prepared by Gross, Schwarz, Laurent, and others all involve approximations. However, the results using these different methods check fairly closely with each other, and the approximations which are involved cannot explain the gross differences between calculated and measured resistances.

The practical significance of the discrepancy was that nearly all the substations were safe from the ground gradient standpoint by a wide margin, regardless of which figure was correct. One or two, however, were apparently safe if we could believe the resistance figure correct and the resistivity figure therefore much too high; modification of the grid to provide a finer mesh might be desirable if it were the resistivity measurement which was representative of actual conditions.

Fortunately, Mr. Kinyon's work has done much to explain the discrepancies. Nevertheless, the matter of what resistivity measurements to take, and which test figure to use in calculations, is likely to remain one of the biggest elements of uncertainty when the designer attempts to apply the principles set forth in reference 1 of the paper and in the proposed AIEE guide.

Additional work still needs to be done, and we hope that Mr. Parker's subcommittee can provide us with further guidance.

The Winding Capacitances in Magnetic Amplifiers

INGE JOHANSEN
ASSOCIATE MEMBER AIEE

THE INFLUENCE of winding capacitance upon the performance of magnetic amplifiers is a matter of some concern. Although no systematic investigation has appeared in the literature, it is generally accepted in the art that such capacitances may be responsible for certain peculiarities in the transfer characteristic of the self-saturating magnetic amplifier. In this paper a lumped equivalent winding capacitance, connected across the winding terminals, is defined. An approximate, empirical formula is given for it for toroidal core windings. The concept of this equivalent capacitance is shown to be of some value in explaining the experimentally observed behavior of certain types of magnetic amplifiers.

The Equivalent Winding Capacitance

It is not obvious that the distributed capacitances of a winding can be replaced by an equivalent, lumped capacitance. In most cases the capacitance between turns within one winding is the most important term. The capacitances between different windings and from winding to ground or to the core will be treated later.

To start the analysis, a winding with uniform distributed turn-to-turn capacitance is considered. In Fig. 1 (A) each turn has the same capacitance to its two adjacent turns, but no capacitive interaction with other turns. When a voltage is induced in such a core winding, it is obvious that both the inductive network and the capacitive network separately will give the same linear voltage distribution along the winding. The networks can then be separated as shown in Fig. 1(B). This is easily justified by the fact that there is no voltage difference between the points A' and A'' and it is accordingly of no importance whether these points are connected or not. It is then possible to define a lumped capacitance that is equivalent to this idealized internal winding capacitance. This statement is valid whenever the capacitive network yields a linear voltage distribution.

Actually, the capacitive network of toroidal windings is not so simple, and it is unlikely to give a linear voltage distribution in itself. However, in magnetic amplifiers, the voltage distribution is dictated essentially by the magnetic flux, which approximately induces the same voltage in all turns. The winding capacitance, important as it may be, is a second order effect as far as the voltage distribution is concerned, and the assumption of a linear voltage distribution remains accordingly acceptable for further purposes in this paper. The equivalent winding capacitance for such a voltage distribution is defined here as a lumped capacitance across the terminals of the winding, affecting the core mmf (magnetomotive force) in an identical way as the actual distributed turn-to-turn capacitances do.

In Fig. 2 some of the turn-to-turn capacitances are indicated. Between the turns n and m there is a capacitance $C_{n,m}$ and a voltage $e(n-m)/N$, where N is the total number of turns, and e is the voltage across the terminals. If the terminal voltage varies by an amount Δe in the time interval Δt, the elementary capacitance $C_{n,m}$ causes a contribution to the core mmf, which is expressed by

$$\Delta(\Delta F)=(n-m)C_{n,m}\frac{n-m}{N}\,\frac{\Delta e}{\Delta t} \tag{1}$$

The contribution of the capacitances of all the turns to the turn n is

$$\Delta F=\frac{1}{N}\,\frac{\Delta e}{\Delta t}\sum_{m=1}^{N}(n-m)^2C_{n,m} \tag{2}$$

Finally, the total mmf, F, contributed by the turn-to-turn capacitances of the whole winding, is obtained by extending the same consideratinn to all the turns, that is, summing n from 1 to N. In this process all the turn-to-turn capacitances are counted twice, and the final expression for the mmf is

$$F=\frac{1}{2N}\,\frac{\Delta e}{\Delta t}\sum_{m,n=1}^{N}(n-m)^2C_{n,m} \tag{3}$$

A lumped capacitance, C_{eq}, across the winding terminals will have the same effect if

$$C_{eq}=\frac{1}{2N^2}\sum_{m,n=1}^{N}(n-m)^2C_{n,m} \tag{4}$$

From equation 4 theoretical values of C_{eq} could be obtained if the values of $C_{n,m}$ were known. For common toroidal windings this is not the case. Even if the turns were wound in an orderly way, the capacitance $C_{n,m}$ would be a rather complicated function of wire and core dimensions, and of m and n. In addition, the turns are wound randomly, and therefore equation 4 is not useful in obtaining a quantitative expression. Nevertheless, some interesting conclusions can be drawn, as shown in the following.

As a theoretical example, the case of a constant value for $C_{n,m}$ for all values of n and m is considered first. For this case equation 4 shows that C_{eq} will increase approximately proportional to the number of turns in the winding. As another extreme, the case when each turn has a capacitance only to its two adjacent turns is considered. Here the use of equation 4 shows that C_{eq} is inversely proportional to the number of turns. In the actual case of randomly wound turns, C_{eq} is likely

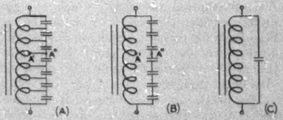


Fig. 1. Winding with uniform distributed capacitance

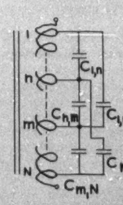


Fig. 2. Elementary capacitances in a winding

Paper 59-776, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted November 3, 1958; made available for printing April 9, 1959.

INGE JOHANSEN is with the Technical University of Norway, Trondheim, Norway.

This work has been supported in part by the U. S. Office of Naval Research, and was performed while the author was a visiting lecturer at the Carnegie Institute of Technology, Pittsburgh, Pa., on a grant from the Technical University of Norway. The author wishes to acknowledge the help of R. W. Roberts of the Westinghouse Electric Corporation in providing many of the wound cores tested, and the guidance and useful suggestions of Prof. L. A. Finzi.

Vol.
Pt.
195

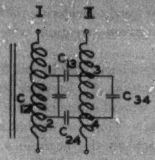


Fig. 3. Fundamental capacitances between two windings on the core

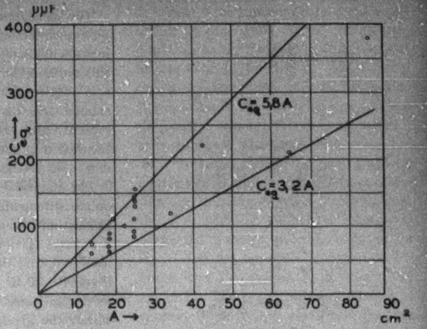


Fig. 4. The measured equivalent capacitance versus the inner surface of the winding

to be between those two theoretical extremes. Hence it is not surprising that actual measurements yield values of C_{eq} that are practically independent of the total number of turns.

It can be shown from equation 4 that C_{eq} is also equivalent to the distributed capacitances in terms of storage of energy. This is a more general property than the one implied by the original definition based on identities of core mmf's.

Equation 4 has been developed under the assumption that the winding has open terminals, but it is recognized that the same equivalent capacitance exists when an impedance is connected across the terminals. In this case the voltage distribution may become less linear, but experiments show that the assumption remains adequate for practical purposes. One interesting conclusion can be drawn at this point. The influence of the winding capacitance upon the core mmf will decrease with decreasing impedance connected at its terminals.

The foregoing discussion has been concerned with cores with one winding only. In practical magnetic amplifier circuitry, the influence of the other windings has to be considered. In Fig. 3 there are two windings on the core and the fundamental winding capacitances are shown. The presence of the second winding has influence in more than one way. To some extent it increases the turn-to-turn capacitances of the original winding, I of Fig. 3, in the same way as a metallic screen would do. The situation is complicated further by the fact that a voltage change in the windings will cause a current to flow through parts of both windings over the capacitances C_{13} and C_{24}. The mmf which results is due to currents in both of the windings. Thus, it is not enough to ascribe an equivalent capacitance to each winding, but the interwinding capacitances should also be considered. For the sake of obtaining useful results the analysis is limited to the case where winding I has many more turns than winding II. In that case the mmf contribution due to the current between 3 and 4, i.e., the part of the current flowing over C_{13} and C_{24}, can be neglected. The capacitance between two turns, C_{34}, will next be considered. Let N_1 be the number of turns between 1 and 2, and N_2 the number of turns between 3 and 4. The capacitance C_{34} will appear to be $k\ (N_2/N_1)^2C_{34}$ in parallel with C_{12}, as it will be transformed to winding I with the square of the ratio of the number of turns. The factor k depends on the mutual inductance between the windings. With the restriction $N_2<<N_1$, the term $k(N_2/N_1)^2C_{34}$ may be neglected in comparison to C_{12}. Thus, the main influence of winding II is the same as that of a metallic screen, and the equivalent capacitance of winding I depends only on its own design and on the location of the second winding. It should also be clear that the measured capacitance of a winding with few turns depends more on the other windings on the core than on its own design.

In the following only cases where $N_2<<N_1$ are considered. In the cases when the winding capacitance is important this will very often be true and the restriction will not limit the practical value of the derived results.

It will now be clearly shown that the winding with many turns is the important one. The empirical formula equation 6, presented later, shows that the value of C_{eq} for a given core is fairly independent of the number of turns. Ascribing the same capacitance C to two different windings 1 and 2 on the same core with number of turns respectively N_1 and N_2, it is easily shown that the contribution to the mmf from the capacitances is respectively CN_1^2de'/dt and CN_2^2de'/dt, where e' is the voltage per turn. Thus the quantity CN^2 is the significant one, and only the winding with the larger number of turns has to be considered.

A simple criterion can be suggested at this point to determine whether the winding capacitance is important or not in a practical design. Evidently this is the case when the mmf caused by C_{eq} is of the same order of magnitude as the change of signal ampere-turns, $N\Delta i_c$, required to control the amplifier over its whole range of operation. In order to make a rough estimate of the mmf caused by C_{eq}, it is evaluated as the product of the turns, N, times the current that flows in C_{eq} with the gate-voltage impressed on it. This implies that this current is discharged through the winding and the mmf thus produced will be added to the control-ampere turns. If the frequency is ω and the amplitude of the voltage per turn is E, then the maximum value of the capacitive ampere-turns is ωCN^2E. If the ratio $\omega CN^2E/N\Delta i_c$ is about 1 or larger, the consideration of the influence of the winding capacitance is justified.

Now it is noticed that the quantity $CE/\omega N\Delta i_c$ can be regarded as fairly constant for a given core, and thus

$$\frac{\omega CN^2E}{N\Delta i_c}=\frac{CE}{\omega N\Delta i_c}\ (\omega N)^2=K\cdot(\omega N)^2 \qquad (5)$$

This shows that the influence of winding capacitance increases roughly with the square of the product of the gate-voltage frequency and the number of turns in a given circuit.

Hitherto, the analysis has been concerned with the "internal" capacitance. The presence of other windings has been regarded to be of importance only to the extent that they increase the capacitance in the winding under consideration. However, the interwinding capacitances may also be of importance. This is so because each winding is connected to external networks, and current can flow in a circuit consisting of the interwinding capacitance, the windings themselves, and the external networks. A further discussion of this point is omitted here.

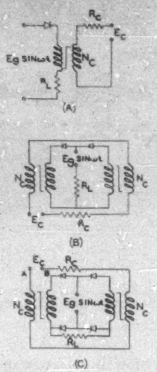


Fig. 5. Amplifier circuits that are examined

A—The half-wave amplifier
B—The doubler circuit
C—The bridge circuit

An Empirical Formula for the Equivalent Winding Capacitance

It has been stated already that an analytic expression for the equivalent winding capacitance is difficult to derive, and that its usefulness is doubtful. Therefore, attempts have been made to obtain an empirical expression. This has met reasonable success, when the accuracy requirements were not too high.

The measurements leading to the empirical expression were performed according to a method given by Terman.[1] With this method the condition of linear voltage distribution is preserved. The capacitance was measured with the core present, and the core was driven into saturation by an additional d-c source or it was short-circuited.

The empirical formula derived is based on measurements on windings with cores having inner diameter from 0.5 inch up to 2 inches. The wire was double enamel insulated, and the range of wire size was from No. 30 to No. 42, Awg (American wire gage) size. The formula is valid within the given range when the other windings have few turns, i.e., up to 20% to 30% of the turns of the winding considered. The radial thicknesses were at least equal to 25% of the available window space. A number of measurements yielded the astonishing simple relation that the capacitance is proportional to the inner surface of the winding, and is fairly independent of the number of turns and of the thickness of the wire. In Fig. 4 the equivalent winding capacitance C_{eq} is plotted versus the inner surface of the winding. It is seen that this capacitance may be expressed as

$$C_{eq} = \alpha \cdot A \ \mu\mu f \text{ (micromicrofarad)} \qquad (6)$$

where A is the inner surface of the winding in cm^2, and α is a constant within the range of 3.2 to 5.8. Hence, the equation has a certain ambiguity of ±30%, but this is quite good considering the fact that the windings in question were wound randomly.

Some comments can be made on the use of equation 6. In some cases the winding is wound randomly, but without overlapping. That means that there is a spacing between the beginning and the end of the winding. This reduces the capacitance somewhat. Either case is within the limit of equation 6, but account of this effect might be taken by choosing the value of the constant α within the over-all range stated, so that an accuracy better than ±30% can be achieved.

It is stated above that the winding has to be of a certain radial thickness. This is especially important when thin wires are used, as, No. 37 to No. 42. In such cases the winding capacitance will increase with decreasing thickness of the winding, and perhaps a better empirical expression should be sought that also accounts for radial thickness. In this work the number of measurements was too small to justify an attempt in this direction, but this effect should also be considered when choosing α within the range. On the other hand the position of the winding (outer or inner winding) has small effect on α.

The surprisingly simple equation 6 needs some analytical justification. The fact that C_{eq} is approximately independent of the number of turns has been discussed already. The fact that the wire size is unimportant can be explained in part by the fact that the ratio between net copper diameter and total diameter varies within narrow limits, between 0.7 and 0.8 for double enamel insulated copper wire in the range considered. Also, it can be understood qualitatively that the capacitance is proportional to some surface of the winding. This surface is equal to the length of a turn multiplied by the mean length of the winding. It is evident that the winding capacitance is proportional to the length of a turn, and an increase in the mean length of the winding will undoubtedly increase the capacitive interaction between the turns, which means an increase in the winding capacitance.

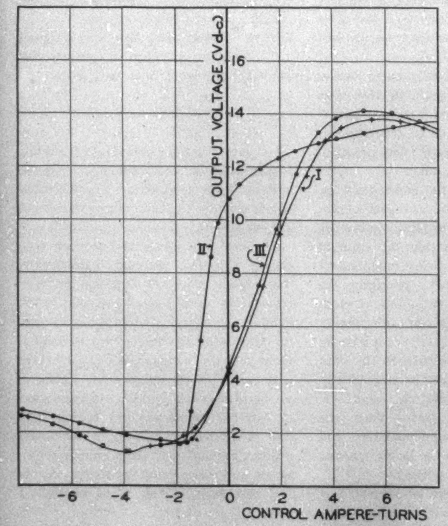


Fig. 6. Transfer characteristics for the half-wave amplifier. $E_g = 35$ volts, $f = 3.6$ kc. I: $N_c = 5{,}000$ turns, $R_c = 5.25$ megohms, $C_{eq} = 100\ \mu\mu f$. II: $N_c = 500$ turns, $R_c = 52.5$ kilohms, $C_{eq} = 100\ \mu\mu f$. III: $N_c = 500$ turns, $R_c = 52.5$ kilohms, $C = 10{,}000\ \mu\mu f$

Special Types of Windings

Only randomly wound windings have been discussed above. In some cases special windings are used to minimize voltages between adjacent layers and also to reduce the winding capacitance. The section winding is a type of such windings. Here the winding is divided in sections that follow progressively along the core, while each section is wound randomly.

Vol.
Pt. 1
1959

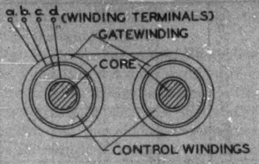


Fig. 7. Cross section of the core and windings

Equation 6 can be used for each section. The resultant capacitance may be obtained as the series connection of the capacitances of the various sections. Hence, it appears that C_{eq} decreases inversely to the square of the number of sections. This is an approximate reasoning, however, because the capacitive interaction between the different sections has not been taken into account. The presence of the core and neighboring windings will mainly provide this coupling, and when the winding is divided into many sections, this will be the dominating factor. For this reason dividing the winding into more than 4 to 6 sections is of no value. If further decrease in the capacitance is desired, it is necessary to increase the distance of this winding from the core and from the other windings, at the same time as the number of sections is increased.

Experiments on Magnetic Amplifiers

The purpose of this section is to show that the distributed winding capacitance can be replaced by an equivalent lumped capacitance, as defined in the first section, in the study of magnetic amplifier behavior. The experiments are based on the fact that two magnetic amplifiers that are identical in other respects, will behave in the same way when the product CN^2 is the same. If the concept of an equivalent capacitance is applicable, this will be the case regardless of whether the winding has distributed capacitance ($C=C_{eq}$) or whether it has little or no capacitance of its own, but is shunted by a capacitor at its terminals. Transfer characteristics of a magnetic amplifier circuit have been measured, where the winding capacitance was significant. This transfer characteristic is later referred to as characteristic I. The equivalent winding capacitance, C_{eq1}, was measured, and the product, $C_{eq1}N_1^2$, was calculated. Then the same amplifier circuit was duplicated with a winding having a fewer number of turns, N_2, so low that its equivalent capacitance ($C_{eq2}=C_{eq1}$) had a negligible effect on the transfer characteristic. The corresponding characteristic is referred to later as transfer characteristic II. Finally an external capacitance, C, was connected across the terminals of this second amplifier, and its value was chosen such that $CN_2^2=C_{eq}N_1^2$. The corresponding transfer characteristic is called characteristic III. Now if the predictions made in this paper are valid, transfer characteristics I and III should be fairly equal.

The influence of the winding capacitance, actual or simulated, upon the amplifier behavior can be seen by comparing I or III to II. In all cases the control winding was chosen as the winding which gives rise to the capacitive effects, since the number of turns could be changed easily without affecting the other parameters in the circuit.

Transfer characteristics I, II, and III were measured for a number of cases. The types of magnetic amplifier circuits that are examined are the half-wave amplifier, the doubler, and the bridge circuit; see Fig. 5. Fig. 6 shows the transfer characteristics I, II, and III for the half-wave amplifier. In this case it was difficult to obtain a truly smooth control current. However, care was taken to use the same value of R_c/N_c^2 in all measurements. It is seen that there is a good agreement between I and III, as predicted, and that the winding capacitance has a marked influence, as shown by the large difference between I and II. The winding capacitance causes a decrease in the slope of the transfer characteristic and a shift to the right. Some care was required in order to obtain good agreement between I and III on Fig. 6. Fig. 7 shows a cross section of the core and indicates the location of the two windings. When the control winding is randomly wound, as in this case, the turns with about the same potential as the terminal *c* will be distributed around the outer surface of the winding, and thus form a screen against the gate winding. When terminal *c* is connected to the grounded terminal of the signal source a capacitive interaction between the two windings is essentially avoided. The way the gate winding is connected is not very important, because the rate of change of the voltage is small compared to that in the control winding. The influence of the interwinding capacitance may be seen by interchanging the terminals *c* and *d*. This influence is by no means unimportant but can in many cases be eliminated.

Fig. 8 shows the transfer characteristics I, II, and III of the bridge circuit with low constraint. To obtain good agreement between I and III, the terminals *A* and *B* in Fig. 5(C) had to correspond to winding terminal *c* in Fig. 7. Good results are obtained since in the low constraint case both terminals have a constant potential in comparison to the induced voltages in the control windings. Transfer characteristic IV in Fig. 8 demonstrates the importance of the interwinding capacitance. In this case the

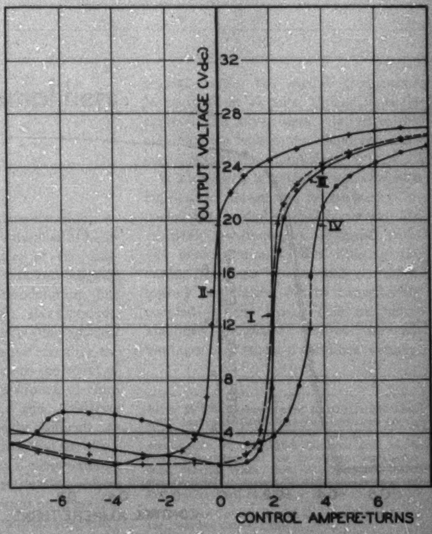


Fig. 8. Transfer characteristics for the bridge circuit. $E_g=35$ volts, $f=3.6$ kc. I: $N_c=5{,}000$ turns; $R_c\simeq 0$; $C_{eq}=100$ $\mu\mu$f. II: $N_c=500$ turns; $R_c\simeq 0$; $C_{eq}=100$ $\mu\mu$f. III: $N_c=500$ turns; $R_c\simeq 0$; $C=10{,}000$ $\mu\mu$f. IV: Identical to I, but the terminals on the control winding are interchanged

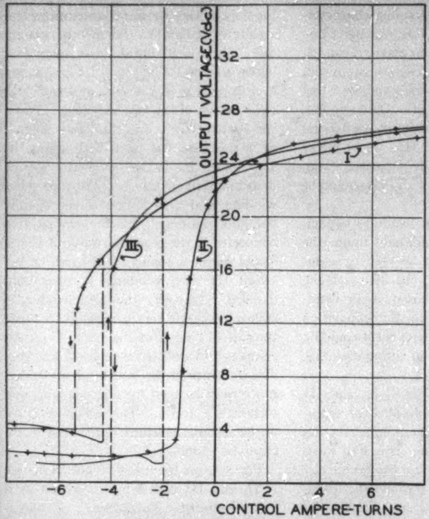


Fig. 9. Transfer characteristics of the doubler circuit. The same data as for Fig. 6

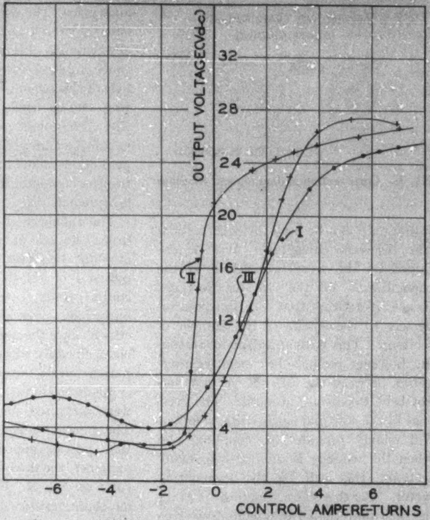


Fig. 10. Transfer characteristics of the bridge circuit. The same data as for Fig. 6

terminals were interchanged for both control windings. The transfer characteristics of the doubler and the center tap are quite similar to that of bridge circuit for the low constraint case.

Figs. 9 and 10 show the transfer characteristics I, II, and III of the doubler and of the bridge circuit for a high constraint case. In all cases R_c/N_c^2 has the same value. The agreement between I and III is not as good as before, because it is impossible to avoid capacitance interaction between the windings. Only one of the terminals of the signal source can have a constant potential in the high constraint case, and to avoid interwinding capacitive interaction a constant potential on each of the control windings would be required. However, in spite of the influence of interwinding capacitances, it seems that the internal winding capacitance is basically the most important. Figs. 9 and 10 show that the characteristics I and III have the same general behavior, although the quantitative agreement is not good. The center-tap circuit

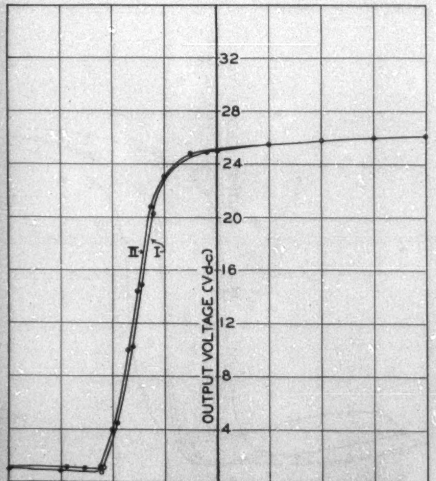


Fig. 11. Transfer characteristics for a doubler circuit where the control winding links both cores. I: $N_c=30{,}000$ turns, $R_c=0$, $C_{eq}=120$ $\mu\mu f$. II: $N_c=1{,}000$ turns, $R_c=0$, $C_{eq}=120$ $\mu\mu f$

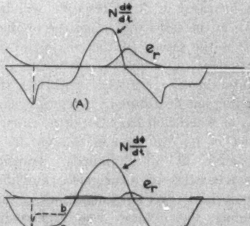


Fig. 12. The load voltage e_r and the core voltage $Nd\phi/dt$ in a half-wave amplifier

A—No influence from the winding capacitance

B—The winding capacitance has an important influence has a transfer characteristic similar to that of the bridge circuit.

Fig. 11 shows the transfer characteristics I and II of a doubler circuit with a control winding linking both cores. The characteristics in this case are almost identical, and the winding capacitances have accordingly no influence. This is in agreement with the predictions made since the capacitance is connected across the signal source directly. Experiments show that the winding capacitance has no influence regardless of the impedance of the control circuit. The bridge, the doubler, and the center-tap circuits behave almost identically, provided the number of control turns is large.

No experimentation has been performed to check on the validity of the "equivalent capacitance" concept during transients. Really, the steady state in itself is a succession of transients and the use of the equivalent capacitance, at least in "slow transients," seems justifiable.

It is not within the scope of this paper to describe why the winding capacitance has an influence on the behavior of magnetic amplifier circuits. Each circuit has several modes of operation. Unsymmetrical and subharmonic modes have been found which are not easily understood. Only the operation of the half-wave amplifier is examined in the following.

Fig. 12 shows the voltage e_r across the load and the induced voltage $Nd\phi/dt$. In Fig. 12(A) the winding capacitance is of minor importance, while it has a larger influence in Fig. 12(B). In case *a* the behavior can be understood by the description in reference 2. For case *b* the analysis is started as the flux is descending from positive saturation. Limiting the case to resistive loads, the current in the gate-winding will be positive, and the voltage induced in the winding is slightly larger than the impressed voltage. Up to point *a*, where the gate-voltage is equal to the induced voltage in the gate winding, the winding capacitance is charged negatively from the gate voltage. At point *a* the rectifier blocks. Instead of following the path *ab*, which indicates the induced voltage in the core under constant mmf, as in Fig. 12(A), the core voltage will follow *ac*, as the winding capacitance will discharge and supply the core with additional negative mmf. The area *abc* can simply be regarded as the contribution of the winding capacitance to the reset of the core. It is noticed that the winding capacitance in this case causes a larger reset. The contribution from the capacitance is, however, dependent on the firing angle and of the induced core voltage under the corresponding constant mmf. A shift to the right and a change in the slope of the transfer characteristic, as observed in Fig. 6, may then be understood.

This explanation gives the key to the understanding of the bridge and the centertap circuit. The doubler behaves in an entirely different way, due to the unblocking of the rectifiers during the reset period. In amplifier circuits with low constraint the foregoing reasoning is not valid. Here it is surprising that the slope is almost constant regardless of the winding capacitance, and the only influence is a shift in the transfer characteristic. No attempt has been made to explain this fact. It suggests, however, that a lumped capacitance across one of the windings could replace a separate bias winding.

Conclusions

In this paper a lumped equivalent winding capacitance C_{eq} has been defined. This definition has been limited to cases where the winding under consideration has a much larger number of turns than any other winding on the core. An empirical expression has been proposed for the winding capacitance which yields results accurate within ±30% or better.

Experiments show that in a number of cases the distributed capacitance has the same effect as an external shunt capacitance at the winding terminals. This enhances the practical usefulness of the concept of C_{eq}.

References

1. Radio Engineers' Handbook (book); F. E. Terman. McGraw-Hill Book Company, Inc., New York, N. Y., first edition, 1943, pp. 84–85, 922–23.

2. Dynamic Core Behavior and Magnetic Amplifier Performance, L. A. Finzi, D. L. Critchlow. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, May 1957, pp. 229–40.

Capacitively Coupled Magnetic Amplifiers

H. WILLIAM COLLINS
ASSOCIATE MEMBER AIEE

AMPLIFICATION of a-c signals with magnetic amplifiers is generally more difficult than d-c amplification. The conventional full-wave circuits, such as the doubler circuit and the bridge circuit, can yield high performance only by demodulating a-c signals and amplifying the resultant direct current. These circuits remain, in effect, d-c amplifiers and suffer from several shortcomings. For example, the sensitivity is limited by the forward voltage drops of the demodulating elements, the output is affected by d-c drifts of the control characteristics, and the component duplication and ineffeciencies of push-pull circuitry must be used to obtain phase sensitivity. The work of Ramey has led to the development in recent years of a number of high-speed magnetic amplifiers of the half-wave type. These circuits can amplify a-c signals without demodulation, but they also have certain shortcomings. The power gain per stage is generally less than the gain of full-wave circuits, the amplifiers remain subject to the effect of zero drift caused by unbalance of component characteristics, and low-level sensitivity is particularly difficult to obtain. Geyger has shown that the effect of amplifier zero drift on a-c servomotors (or other frequency discriminating output elements) can be eliminated by operating half-wave circuits at a carrier frequency several times the signal frequency.[1] None of these circuits, though, entirely achieves the advantages relative to d-c amplification techniques which are associated with true a-c amplifiers.

If a magnetic amplifier is operated at a high-carrier frequency, its frequency response is extended so that it can amplify relatively low-frequency a-c signals without demodulation. After filtering the carrier-frequency pulses, the signal-frequency modulation can be capacitiviely coupled to the load. This magnetic-amplifier technique realizes the advantages of a true a-c amplifier, namely:

Paper 59-785, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted August 31, 1958; made available for printing April 14, 1959.

H. William Collins is with Crydom Laboratories, Inc., Garden Grove, Calif.

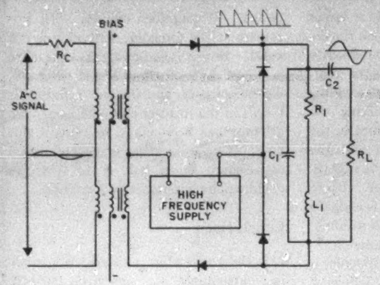


Fig. 1 (left). Capacitively coupled magnetic amplifier

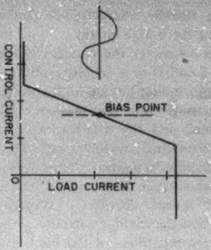


Fig. 2 (right). Idealized magnetic-amplifier control characteristic

freedom from the effect of d-c drift, consequent increased sensitivity, inherent phase reversibility, and accurate reproduction of the input waveform.

Capacitively Coupled Circuit

A capacitively coupled magnetic amplifier using a full-wave bridge crcuit is shown in Fig. 1. The high-frequency power can be supplied by a frequency multiplier[2] or a transistor inverter.[3] The carrier frequency should be at least 10 times the maximum signal frequency that is to be passed; otherwise, the amplifier gain is unduly reduced by the residual delay of the amplifier and it is difficult to filter the carrier-frequency pulses without attenuating the signal-frequency power. Magnetic amplifiers can be successfully operated in the radio-frequency range.[4] The limitation on the pass band of a high-frequency magnetic amplifier is, in practice, set by the maximum operating frequency of a satisfactory power supply. Both frequency multipliers and transistor inverters have ruggedness and reliability commensurate with magnetic amplifiers and are suitable for this application. Appreciable power can be obtained from these devices at a maximum operating frequency of about 20 kc.

A capacitively coupled amplifier is biased at approximately one-half output (see Fig. 2) and the incoming a-c signal modulates the average output about this quiescent level. Capacitor C_1 filters the carrier-frequency pulses, and a smooth d-c having a signal-frequency modulation is produced across the output. The quiescent d-c passes through R_1 and L_1. This path should be a rather low impedance to d-c and, for maximum output power, a high impedance to signal-frequency power as provided by inductor L_1. Capacitor C_2 couples the signal-frequency power to the load and blocks the d-c component. An amplified replica of the a-c input signal is thereby produced in the load.

Except for the use of a carrier-frequency power supply, the circuit operation is analogous to the fundamental a-c vacuum-tube technique. The advantages of capacitively coupled magnetic amplifiers relative to the standard magnetic circuits are listed as:

1. Drift-free. The effect of zero-signal drift caused by component variations is eliminated, and a major practical limitation on magnetic-amplifier sensitivity therefore removed.

2. No demodulator. A-c signals can be amplified without demodulation by a high-gain full-wave circuit. Not only is a critical circuit element eliminated, but the corresponding forward voltage drops which limit sensitivity are also removed.

3. Signal sensitivity. A-c signals can be amplified which are several orders of magnitude smaller than possible with standard magnetic-amplifier techniques.

4. Phase sensitive. The 2-core circuitry is inherently phase sensitive. The duplication of components in push-pull circuitry is not required nor is there a d-c component in the output.

5. High speed. The amplifier begins response within 1/2 cycle of the carrier frequency and reaches full response in a fraction of 1/2 cycle of the signal frequency.

6. High gain. The use of a high-carrier frequency allows a greatly increased figure of merit to be realized.[5] In amplifying low-frequency signals, e.g., 60 cps (cycles per second), many cycles delay of the carrier frequency are tolerable and high power gains per stage are obtainable.

7. High stability. Because of constancy of the differential permeability and inherent voltage stability, the circuitry has excellent gain stability in addition to freedom from zero drift.

8. Linearity. Excellent linearity is achieved.

9. Small size. A high-supply frequency results in greatly reduced component size and weight for a given power rating.

10. High fidelity. Except for a very slight ripple, waveforms within the pass band of the amplifier are accurately reproduced in the output.

Nomenclature

A_p = power gain as defined by equation 2
A_p' = power gain at the amplifier-break frequency
E_c = a-c control voltage, rms volts
E_0 = a-c output voltage, rms volts
I_c = a-c control current, rms amperes
K = gain constant, volts per ampere
K_v = d-c voltage gain
L_c = effective inductance of control winding, henrys
M = figure of merit in terms of power gain per second of response time
R_c = total control circuit resistance, ohms
R_L = load resistance, ohms
T_1 = time constant of coupling capacitor and load resistance $= R_L C_2$
T_2 = time constant of magnetic amplifier $= L_c/R_c$
T_3 = time constant of filter capacitor and load resistance $= R_L C_1$
w_s = signal frequency, radians per second
w_s' = signal frequency at amplifier break point
X_1 = effective control-circuit reactance, ohms

Design Considerations

The maximum average a-c output power of a capacitively coupled magnetic amplifier is much less than the maximum instantaneous d-c power which the amplifier must supply. Note from Fig. 2 that the peak a-c output voltage is one-half the maximum d-c output voltage and that the maximum undistorted sinusoidal voltage has an rms value of $\sqrt{2}/4 = 0.353$ times the peak d-c voltage. The nonlinearities at the ends of the control characteristic limit the maximum excursion which will yield an undistorted output. The net result is that the maximum average a-c power delivered by a capacitively coupled amplifier operating in the linear region is less than 1/10 the instantaneous d-c power which the amplifier must supply at peak output current. The control power to produce the reduced average flux excursions, though, is similarly decreased and the power gain of the capacitively coupled amplifier remains the same as that obtainable with the d-c amplifier of equivalent response time. The power obtainable from a high-frequency supply is, in practice, limited, and therefore it is difficult to construct capacitively coupled amplifiers with high output power. Since capacitively coupled magnetic amplifiers are characterized by limited output power combined with excellent stability and a-c sensitivity, this technique is most applicable for low-level amplification. Major applications appear to be servo preamplifiers, strain-gage amplifiers, and, by the use of a chopper, low-level d-c amplifiers.

Referring to Fig. 1, capacitors C_1 and C_2 affect the frequency response and power gain of the amplifier. Capacitor C_1 should be large enough so that there is negligible attenuation to the lowest a-c signal frequency. Capacitor C_1 must remain small enough so there is negligible attenuation of the highest signal frequency. The response of a single-stage capacitively coupled amplifier can be expressed by the following transfer function:

$$E_o/E_c=\frac{K_v\,\dfrac{jwT_1}{1+jwT_1}}{(1+jwT_2)(1+jwT_3)} \qquad (1)$$

The break point on an attenuation versus frequency plot caused by T_1 (the coupling capacitor) should be below the lowest signal frequency of the desired amplifier pass band. The break resulting from T_3 (the filter capacitor) should be above the highest desired signal frequency. As the signal frequency approaches the carrier frequency, T_3 must become relatively smaller. As a result less filtering of the carrier pulses is achieved and an undesirably high "hum" level is pesent. In many applications where a frequency-discriminating-output device is driven (such as another magnetic-amplifier stage or a servomotor) the presence of this ripple, which is twice the carrier frequency, is not detrimental. In other applications though, such as driving a meter, this spurious output appears as a zero-signal error. At high signal frequencies it may be necessary to use an additional low-pass filter stage (such as series inductor in the output) to obtain increased discrimination between the signal frequency and the carrier ripple. In practice the carrier ripple can easily be held to less than 1% of full output with signal frequencies as great as 0.1 of the carrier frequency.

Power Gain and Frequency Response

According to the figure-of-merit concept, the power gain of a full-wave magnetic amplifier varies inversely with response time. In order to evaluate the power gain of a full-wave magnetic amplifier, the effective input impedance must be known. The effective input impedance of a magnetic amplifier operated as a d-c amplifier is simply the total series resistance of the control circuit. When a magnetic amplifier is subjected to a continually varying signal, the input impedance has an inductive as well as a resistive component. A magnetic amplifier with a multicycle delay responds exponentially to a change in control and can be properly represented by a series L_c-R_c circuit. The output voltage of a given amplifier is proportional to a constant times the control current. The control current for a given control voltage is a function of the input impedance.

$$E_o=KI_c=K\,\frac{E_c}{R_c+jw_sL_c}=K\,\frac{E_c/R_c}{1+jw_sT_2} \qquad (2)$$

In this discussion power gain is defined as follows:

$$A_p=\frac{E_o^2/R_L}{E_cI_c} \qquad (3)$$

It will be noted from equation 2 that as R_c is made small relative to w_sL_c, the input impedance becomes purely reactive and the real power consumed by the control circuit approaches zero. This condition is of only academic interest as the control source must still possess a volt-ampere capability which determines the output power. Hence, the realistic method of specifying the required control signal is in terms of volt-amperes ($E_c \times I_c$).

At a given signal frequency, maximum gain is achieved when R_c is small and the control source essentially works directly into the winding inductance. That is, if the input impedance is minimized, the least possible control voltage is required to force the control current which produces a given output. Thus, a given output is produced with minimum volt-ampere signal and maximum power gain. Sufficient copper can easily be placed in the control winding of a magnetic amplifier so that the control-circuit resistance is small relative to the winding reactance (w_sL_c) at power-frequency signals. In practice the control-circuit impedance can be reduced to essentially the value of the reactance and the condition of maximum possible power gain can be obtained. This design condition corresponds to operation of the amplifier at a frequency beyond the break point determined by T_2 in equation 1. As a result of operating beyond the break point, the relative gain (20 log E_o/E_c) is falling at a rate of 20 per decade of frequency. This rate of decrease is shown in the upper curve of Fig. 3. If the control source works directly into the winding reactance, maximum possible power gain is always achieved, but the amplifier obviously does not have a flat-frequency response which is required in many applications.

A flat-frequency response can be obtained by moving the break point ($1/T_2$) beyond the maximum desired signal frequency. This is done by making the control-circuit resistance greater than the reactance at the desired signal frequency. The relationship between gain and signal frequency for various ratios of control-circuit resistance to reactance is shown in Fig. 3. The curves do not consider the attenuation caused by the amplifier filter or coupling capacitors, as it is presumed that their break points can be placed outside of the desired pass band. These curves demonstrate that, in applications where the signal frequency varies, the gain can be held constant at the expense of additional control power.

At signal frequencies where the reactance is the dominant component of the input impedance, the input impedance increases directly with signal frequency. Under these conditions the voltage gain varies in inverse proportion to the signal frequency. Because of increasing input impedance, the power gain likewise varies in inverse proportion to the voltage gain. In other words, the power gain varies as a function of signal frequency directly with

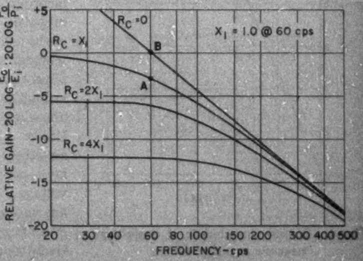


Fig. 3. Relative gain-frequency response as function of control circuit impedance

voltage gain, rather than as the voltage gain squared, as occurs in the case of an amplifier of constant input impedance. In determining the relative power gain from Fig. 3, the relationship 20 log P_o/P_c or 20 log E_o/E_c must be used. For example, if R_c is increased from zero to a value equal to X_c at 60 cps, both the 60-cps voltage and power gain are attenuated by 0.707. It will be noted from Fig. 3 that quite flat frequency response can be obtained without excessive gain attenuation. That is, if the control resistance is made equal to twice the control reactance at a given signal frequency, the gain is flat within 1 db (decibel) to that frequency, yet the power gain is only halved relative to the maximum possible value.

The figure of merit of a magnetic amplifier (M) in terms of power gain per second of response time is a valuable means of relating the power gain and transient response of a full-wave magnetic amplifier. The maximum power gain which can be obtained under sinusoidal a-c control at various signal frequencies can be determined from the figure of merit by the following relationships. Considering only the effect of the amplifier time constant (T_2), it can be seen from equation 1 that the gain is down by a factor of $\sqrt{2}$ at the frequency were $w_sT_2=1$. This frequency corresponds to the break point of the amplifier and is shown at point A in Fig. 3. The signal frequency of a given amplifier at which this break point occurs is

$$w_s'=\frac{1}{T_2} \qquad (4)$$

The steady-state gain is reduced by a factor of $\sqrt{2}$ at the break frequency.

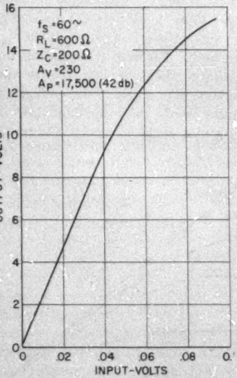


Fig. 4. Control characteristic of single-stage magnetic amplifier, 60 cps

$$A_p'=\frac{MT_2}{\sqrt{2}}=\frac{M}{\sqrt{2}w_s'} \qquad (5)$$

Maximum possible power gain will be obtained when the control resistance is small relative to the reactance at a given signal frequency, and this gain is $\sqrt{2}$ times the gain obtained when the break point occurs at this frequency. Reduction of R_c toward zero corresponds to changing the operating condition of the amplifier from point A to B on the curves of Fig. 3. The amount of increased gain is evident. The maximum possible gain (when $R_c=0$) obtainable at any signal frequency can, therefore, be expressed:

$$\text{Maximum } A_p=\frac{\sqrt{2}M}{\sqrt{2}w_s}=\frac{M}{w_s} \qquad (6)$$

This expression states the limiting case of maximum power gain and applies when w_s is large relative to w_s'.

The power gain at any signal frequency can be expressed as follows in terms of the figure of merit and the time constant.

$$A_p=\frac{M}{w_s'(1+jw_sT_2)}=\frac{MT_2}{(1+jw_sT_2)} \qquad (7)$$

This equation is accurate at a maximum signal frequency of about 0.1 the carrier frequency and over the frequency range where attenuation of capacitors C_1 and C_2 is negligible.

Amplifier Stability

The figure of merit of an audio-frequency magnetic amplifier[5] is in the range from 10^6 to 10^7. Equation 6 indicates that a power gain per stage in amplifying a 60-cps signal can be as high as 26,000 (44 db). The practical power gain which can be obtained from an amplifier is closely related to the stability of the amplifier. Several features indicate that a capacitively coupled magnetic amplifier is basically a very stable device.

Numerous tests have shown that the core gain (differential permeability) of a magnetic material does not display aging and is quite constant with varying ambient temperature. The major change in magnetic core materials is a lateral shift of the control characteristic with ambient temperature variations. Capacitive coupling eliminates the effect of this shift, as long as operation is maintained in the linear region of the control characteristic. Unlike a vacuum tube or a transistor amplifier, the gain of the magnetic amplifier is not a direct function of the supply voltage. Even though the supply voltage varies, a given control signal causes a given flux change. Therefore in a capacitively coupled magnetic amplifier the same modulation around the bias point occurs, and the output voltage remains constant. If the amplifier operates from a transistor inverter, there is an additional reason for stability as the inverter has a constant volt-second output (frequency increases at the same rate as supply voltage) which is ideal for operation of a magnetic device. If an a-c magnetic amplifier is operated from a voltage source and has negligible control-circuit resistance, the gain will be independent of the differential permeability of the core material. This condition occurs because the core gain and the input impedance (reactance) vary in a similar manner. Thus the control current also varies in a manner which maintains a constant output. Stating this another way, the amplifier is voltage controlled and the control signal forces a certain flux excursion which produces a given output, regardless of the characteristics of the core material.

The net stability of a capacitively coupled magnetic amplifier has yet to be fully evaluated. The aforementioned factors indicate, though, that the device may offer much greater stability for a given gain than other amplifying elements Starting the converse, a magnetic amplifier may allow a much higher gain for a given degree of stability.

Single-Stage Amplifier

A capacitively coupled full-wave bridge circuit as shown in Fig. 1 has been constructed. The amplifier operates from a transistor inverter at a 60-volt supply and at a frequency of 10 kc. The system (inverter and bias supply) is driven by a 24-volt d-c supply. It was developed primarily to amplify 60-cps signals. The cores consist of 32 wraps of 1/2-mil *4-79* Mo-permalloy* on 1/2-inch-inside-diameter ceramic bobbins. The load winding has 1,800 turns and the control winding has 350 turns. Germanium diodes and tantalum capacitors are used in the circuit. The maximum d-c output power of the stage is about 4.0 watts. The inverter and amplifier components fit easily into the space of a 3-inch cube.

A control characteristic when amplifying a 60-cps signal is shown in Fig. 4. The small signal linearity of the amplifier is excellent. Although not clearly evident on the curve, the nonlinearity of the amplifier below 8-volt output is too small to separate from the normal error of 1% meters. The control circuit inductance of the amplifier is 0.53 henry which produces a minimum input impedance of 200 ohms at 60 cps. The effective input re-

---

*Trade name.

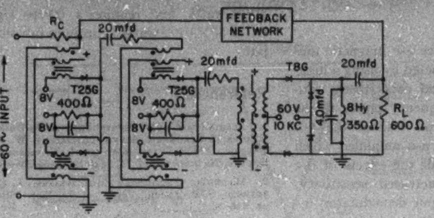


Fig. 5 (above). Circuit diagram of 3-stage amplifier

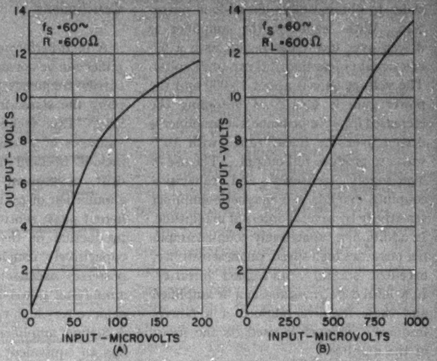


Fig. 6 (right). Control characteristic of 3-stage amplifier, 60 cps

A—Open loop
B—Closed loop

actance has been determined experimentally by increasing R_c until the output is 0.707 of the value that existed with R_c very close to zero. At the 0.707 point R_c equals the effective control-circuit reactance. Working directly into the input reactance, the 60-cps voltage gain is 230 and the power gain, as defined by equation 2, is 17,500 (42 db). The maximum usable a-c output power occurs at about 16 volts and equals 0.4 watt. The amplifier is driven to full output with less than 0.1 volt and 50 μw (microwatts) of signal. About 0.1 volt of 20-kc ripple is present in the output waveform with only a capacitor, C_1, used as a filter. Variations in the 24-volt d-c supply of ±10% cause less than ±2% change in gain.

Multistage Amplifier

Since a capacitively coupled technique eliminates the effect of d-c drift, it is possible to modulate an amplifier over only a small portion of its control characteristic and still maintain excellent stability. Likewise, since there are no rectifiers or other demodulating elements in the control circuit of these amplifiers, drifts in the characteristics of these components do not affect the practical sensitivity which can be realized. The practical sensitivity of any amplifier is eventually limited by the noise level of the amplifier components. By eliminating the effect of d-c drift and the voltage drops of demodulating elements in a magnetic amplifier, the major obstacles which have limited magnetic-amplifier sensitivity are removed. It now becomes possible to approach the true noise level of a magnetic amplifier. The most probable inherent sources of noise in a magnetic amplifier are the Barkhaussen effect in the magnetic material, thermal noise in diodes and resistors, and the effect of stray fields on the magnetic cores.

In order to demonstrate the extreme sensitivity possible with a capacitively coupled magnetic amplifier, the stage discussed in the preceding section has been combined with two low-level stages. A circuit diagram of the complete amplifier is shown in Fig. 5. The low-level stages each use cores consisting of 10 wraps of 1/4-mil *4-79* Mo-permalloy on 1/2-inch-inside-diameter ceramic bobbins and operate at an 8.0-volt level. Great difficulty was met in stabilizing the amplifier with the three stages operating from a common inverter. The loading effect of the output stage on the inverter, as the stage was modulated at a 60-cps rate, caused a spurious signal to be injected into the lower stages and unstable operation resulted. To overcome this difficulty, the power supply consists of a low-power inverter which drives the two low-level stages and also switches a pair of power transistors which supply the power to the output stage. A small amount of d-c negative feedback is used around the low-level stages to obtain a margin of stability. The amplifier is working at a signal level where care must be taken in shielding the components from stray fields. The cores of the low-level stages are mounted in tube shields and the critical leads are also shielded.

A 60-cps control characteristics of the amplifier is shown in Fig. 6(A). The input impedance of the amplifier consists of 130 ohms resistance plus the winding reactance and equals about 200 ohms. The amplifier is driven to the top of its linear region with a 100-μv (microvolt) signal which corresponds to an input power of about 50×10^{-12} watts. The voltage gain is 90,000 and the power gain is 27×10^{8} (95 db).

Fig. 6(A) shows that about 0.4-volt output exist with zero-signal input. This output results from the carrier ripple, voltage pickup in the amplifier, and random noise generated in the first stage. This output is the noise level of the amplifier and, referred to input, is the equivalent 5 μv or 10^{-13} watts of signal. A major portion of the zero-signal output is 120-cps ripple which apparently is coming through the transistor inverter from its d-c power supply. The remainder of this output is random noise which mostly occurs in the frequency range from 10 to 100 cps where the gain of the amplifier is near maximum.

Negative feedback can be used around the amplifier and the well-recognized advantages of increased stability, increased linearity, increased input impedance, and reduced noise, are obtained. Fig. 6(B) shows the control characteristic of the amplifier when operating with approximately 16 db of feedback. The open-loop and closed-loop frequency-response curves are shown in Fig. 7. Because of the large amount of phase shift of the open-loop amplifier, it is necessary to use a compensating feedback network in

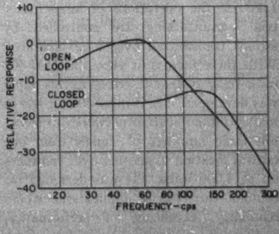


Fig. 7. Frequency response of 3-stage amplifier

order to obtain stable closed-loop operation. With feedback the amplifier is driven over the linear region with a 60-cps signal of less than 1.0 mv (millivolts). The voltage gain is about 15,000 and the power gain is 4×10^8. Considering the increased input impedance, the amplifier is driven over the linear range with 10^{-9} watts to produce full output. The use of high-supply frequency and capacitive coupling extends a-c magnetic-amplifier sensitivity by several orders of magnitude. In addition, a capacitively coupled amplifier possesses high speed, phase sensitivity, excellent stability, and good linearity. Low-level d-c signals also can be amplified and these features retained if a power-frequency chopper is used on the input of the amplifier.

Conclusions

A high-performance a-c magnetic amplifier can be constructed by using a high-carrier frequency and capacitively coupling the signal-frequency power to the load. This technique realizes the advantages of a true a-c amplifier which include freedom from the effects of d-c drift, consequent increased sensitivity, elimination of an input demodulator, inherent phase sensitivity, and accurate reproduction of the input waveform. A capacitively coupled magnetic amplifier possesses excellent stability, and when amplifying power-frequency signals, can achieve very high voltage and power gains. A 3-stage amplifier has been built which has an equivalent noise level of 5μv, or 10^{-13} watts, and requires a 60-cps signal of 50×10^{-12} watts to produce full output.

References

1. Elimination of Asymmetry Zero-Drift Errors in Magnetic Servo Amplifiers, W. A. Geyger. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, May 1955, pp. 239–45.

2. Magnetic Frequency Multipliers, L. J. Johnson, S. E. Rauch. *Ibid.*, vol. 73, Nov. 1954, pp. 448–52.

3. A Switching Transistor D-C to A-C Converter Having an Output Frequency Proportional to the D-C Input Voltage, G. H. Royer. *Ibid.*, vol. 74, July 1955, pp. 322–26.

4. High-Frequency Operation of Self-Saturating Magnetic Amplifiers, H. William Collins. *Ibid.*, Sept., pp. 500–05.

5. Application of High Frequency Magnetic Amplifiers, H. William Collins. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 11, 1955, pp. 136–44.

A New High-Capacity Microwave Relay System

C. G. ARNOLD
NONMEMBER AIEE

V. E. ISAAC
NONMEMBER AIEE

H. R. MATHWICH
NONMEMBER AIEE

R. F. PRIVETT
NONMEMBER AIEE

L. E. THOMPSON
NONMEMBER AIEE

THE INCREASE in demand for communication service throughout the world has exceeded the rate of growth in population and the tremendous expansion of industrial activities.[1] Both civilian and military requirements have grown beyond the expectations of those associated with the communication field. Along with this need for increased capacities in communication networks, the need has developed for more reliable, higher quality communication service. Military operations and the defense of a country are dependent on the ability to communicate rapidly and accurately. Likewise the operations of large industrial organizations are controlled by signals transmitted over long distances. With these factors in view the *MM-600* radio relay system described in this paper was developed. It is intended for use in major communication networks to transmit large numbers of signals from one central location to another over distances up to 4,000 miles. It was designed for common-carrier-type applications where reliability, excellence of performance, and ease of maintenance are of paramount importance. New components such as travelling-wave tubes and solid state devices plus new concepts such as computer techniques have been integrated into proven methods to fulfill these requirements.

This radio system is a line-of-sight system operating in the radio frequency band 1,700 to 2,300 mc (megacycles). It utilizes frequency modulation and employs heterodyne repeaters at an average of 30-mile intervals to compensate for propagation loss. It is designed for satisfactory transmission of either a large number of message channels, of 4 kc nominal bandwidth, or one television signal on a single r-f (radio frequency) channel.

Facilities are provided for operating up to 12 r-f channels simultaneously. The r-f allocation plans, which have been adopted, follow recommendations of the International Radio Consultative Committee (CCIR); however, the flexibility of the system design permits deviations from these standard allocations for specific customer requirements. A typical plan is shown in Fig. 1. In general, because of the need for extremely reliable communication service, it is not desirable to operate a broadband channel without standby facilities. In the *MM-600* design, frequency diversity techniques are utilized to provide automatic protection against either equipment failure or selective fading in the propagation medium. Consequently, it is anticipated that in most applications one or more radio channels in each direction of transmission will be reserved for diversity uses.[2]

System Objectives

The capacity of each broadband channel is nominally 600 message channels or alternately one monochrome or color television signal. A baseband width of 2.5 or 3 mc, depending on the type of frequency division multiplex used, is required in the message application and a maximum of 6 mc is required for television of 625-line standard.

The bandwidth of the equipment for each r-f channel is 30 mc to the half-power points. It is anticipated that this large available FM (frequency modulated) signal bandwidth will permit increased traffic capacities after operating

---

Paper 59-874, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 20, 1959; made available for printing April 29, 1959.

All the authors are with the Radio Corporation of America, C. G. Arnold in New York, N. Y.; V. E. Isaac with RCA Victor, Ltd., in Montreal, Que., Canada; H. R. Mathwich, R. F. Privett, and L. E. Thompson at Camden, N. J.

The integrated efforts of many people were necessary to the success of this development program. The authors, who were responsible for system planning and testing, wish to acknowledge the assistance of their co-workers, particularly those who were engaged in unit development.

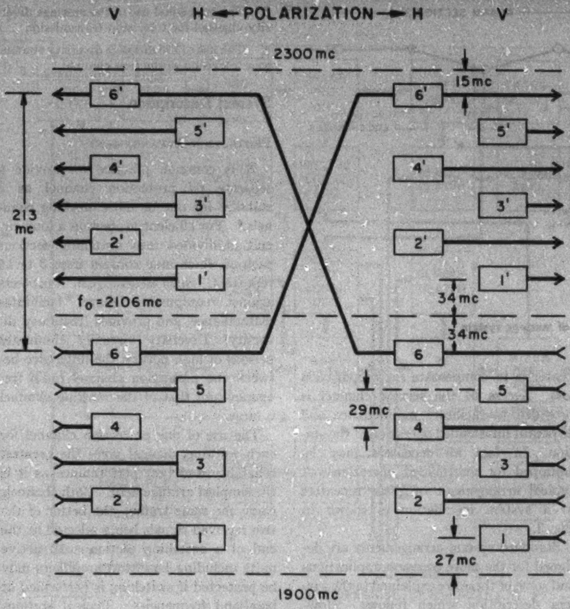


Fig. 1. R-f allocation plan

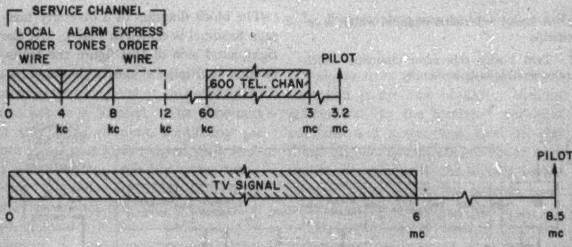


Fig. 2. Baseband allocation plan

experience has been gained and necessary refinements in design are made. The baseband frequency allocation plan is shown in Fig. 2. It will be noted that service channel signals, which are used in maintaining the radio system, can be transmitted on the same channel as the message traffic. This provision obviates the necessity for providing separate communication facilities such as wire lines or another radio system to handle this service.[3]

In establishing the system and unit design requirements for the *MM-600* system, all common kinds of message signals as well as television were considered. Satisfactory performance is realized for ordinary voice, telegraph, telephoto, data, and program. CCIR performance standards have been used as over-all system design requirements. These standards are defined in terms of a reference circuit 1,550 miles long. The basic system parameters were established primarily on message requirements which are, in general, more limiting than the television requirements. Over the reference circuit the allowable signal-to-noise ratio is 53 db (decibels) mean value in the busy hours in the absence of fading; 51 db mean value in any hour. These are the ratios of test tone level, 0 dbm (power in decibels referred to 1 milliwatt), to weighted noise (random circuit noise plus intermodulation noise).

The first value establishes primarily the linearity of the equipment and the second fixes the level of random circuit noise. The latter also requires the adequate provision be made for selective fading.

The basic characteristics of the system are summarized in the Appendix. By using the necessary values in this table, one may compute the random circuit noise in the worst channel of a 600-channel system at 2.54-mc baseband frequency. The ratio with weighting in a 1,550-mile system is 61.5 db. In this computation free space transmission is assumed and no advantage is taken of the fact that in a practical case a given telephone channel need not be transmitted throughout the full length of a system at the top edge of the baseband spectrum. If a 5-db degradation is assumed for the difference between free space loss and mean propagation loss and 2 db is assumed to be gained by shifting the location of the channel lower in the baseband, a net signal-to-noise ratio of 58.5 db is obtained. Further improvement may be achieved, if necessary, by the use of emphasis networks which increase the channel frequency deviation at high baseband frequencies.

Nonlinearity in a radio system results in the generation of modulation products from one or more message channel signals.[4] These products, which are primarily harmonics, and sum and difference signals, create interferences which are noise-like in character.[5] Because of the large number of signals transmitted, the linearity requirements of the system are very severe and many precautions have been taken to insure adequate performance. Reflections in the feeder system are minimized by the use of a well-matched antenna and by isolators at the radio bays. Modulators and discriminators which are linear over a wide band are employed, and delay equalizers are used to compensate for the phase distortion inherent in the r-f and i-f (intermediate-frequency) networks.

The major problems encountered in the transmission of video signals arise from the same sources as those in the message application. The signal-to-noise performance for a standard 525-line television signal over a circuit of reference length is 50 db. This is the ratio of video signal peak-to-peak (excluding synchronizing

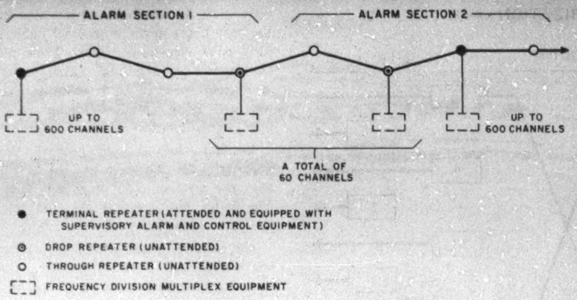

ALARM SECTION 1
ALARM SECTION 2
UP TO 600 CHANNELS
UP TO 600 CHANNELS
A TOTAL OF 60 CHANNELS
● TERMINAL REPEATER (ATTENDED AND EQUIPPED WITH SUPERVISORY ALARM AND CONTROL EQUIPMENT)
DROP REPEATER (UNATTENDED)
○ THROUGH REPEATER (UNATTENDED)
FREQUENCY DIVISION MULTIPLEX EQUIPMENT

Fig. 3. Typical section of message system

signals) to rms random circuit noise. The energy distribution of this noise is essentially triangular: 6 db per frequency octave. Acceptable performance for this kind of interference is 41 db.

System nonlinearity is a major problem in television transmission for color signals for which the differential gain and phase characteristics of the system at the color subcarrier frequency, 3.58 mc, are critical. Emphasis networks are used to alleviate this problem.

Other important system requirements, such as those specifying baseband amplitude response and delay response, gain stability and allowable circuit interruptions, have been realized by the use of appropriate circuit designs and premium components. The equipment is designed to operate through a wide range of environmental conditions.

System Application

The *MM-600* system includes three basic types of repeater:

1. Terminal repeater
2. Drop repeater
3. Through repeater

A terminal repeater would normally be located near a communication center where access is required to a large number of the transmitted message channels. At this repeater all signals are demodulated to baseband frequencies. It is anticipated that in many installations there will be a demand for access to a small number of message channels along a radio path. To fulfill this need without a terminal repeater, the drop repeater is provided. Any number of message channels may be dropped and up to 60 channels inserted. The drop repeater therefore adds to the flexibility of the system.

The only function of the through repeater is to compensate for propagation loss. Access to the service channel is provided to facilitate maintenance and to permit unattended operation of the station. In fact all repeaters may be equipped for unattended operation. A typical arrangement of these repeaters in a system installation is shown in Fig. 3.

Standard system arrangements are designed for the more common applications and some of these are explained in the system description which follows. However, other traffic requirements can be satisfied with only minor modifications in the standard design. Some of these are:

1. A 1-way television channel which is reversible
2. Two 1-way television channels which are reversible independently
3. The occasional use of the message diversity channel for television transmission
4. The use of the message diversity channel as a television protection channel

System Description

Protection Arrangements

It is common practice to provide a separate r-f protection channel as a standby for one or more working channels.[6] For efficient protection a long circuit is divided into switching sections each of which may contain from 3 to 15 repeaters. Such an arrangement protects against equipment failure, facilitates maintenance, and provides frequency diversity. Diversity virtually eliminates periods of high noise if the difference between the protection channel radio frequency and that of the working channel is large.

The use of one protection channel for each working channel gives the greatest reliability and best performance as it is the simplest arrangement. Both channels carry the same traffic; the better of the two received signals being selected at the end of a switching section. All active units including baseband amplifiers may be protected if switching is performed at baseband frequencies. Typical stations using this principle will now be described.

A Radio Terminal

The block diagram of a diversity message terminal is shown in Fig. 4. At the right hand side of this figure the transmitting multiplex signal is shown enter-

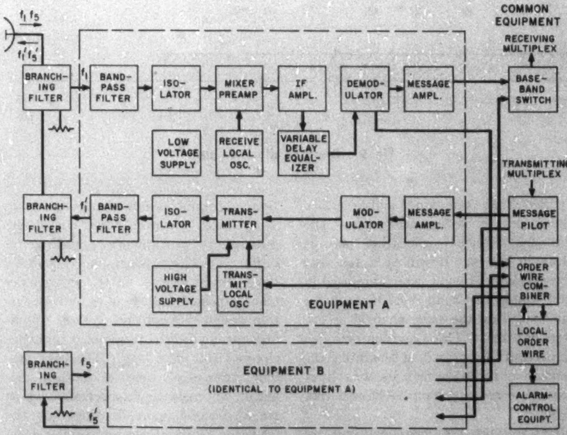

$f_1\ f_5$
$f_1'\ f_5'$
COMMON EQUIPMENT
RECEIVING MULTIPLEX
BRANCHING FILTER
f_1
BAND-PASS FILTER
ISOLATOR
MIXER PREAMP
IF AMPL.
DEMODULATOR
MESSAGE AMPL.
BASE-BAND SWITCH
LOW VOLTAGE SUPPLY
RECEIVE LOCAL OSC.
VARIABLE DELAY EQUALIZER
TRANSMITTING MULTIPLEX
BRANCHING FILTER
f_1'
BAND-PASS FILTER
ISOLATOR
TRANSMITTER
MODULATOR
MESSAGE AMPL.
MESSAGE PILOT
HIGH VOLTAGE SUPPLY
TRANSMIT LOCAL OSC
EQUIPMENT A
ORDER WIRE COMBINER
LOCAL ORDER WIRE
BRANCHING FILTER
f_5
f_5'
EQUIPMENT B
(IDENTICAL TO EQUIPMENT A)
ALARM-CONTROL EQUIPT.

Fig. 4. Diversity radio terminal for message traffic

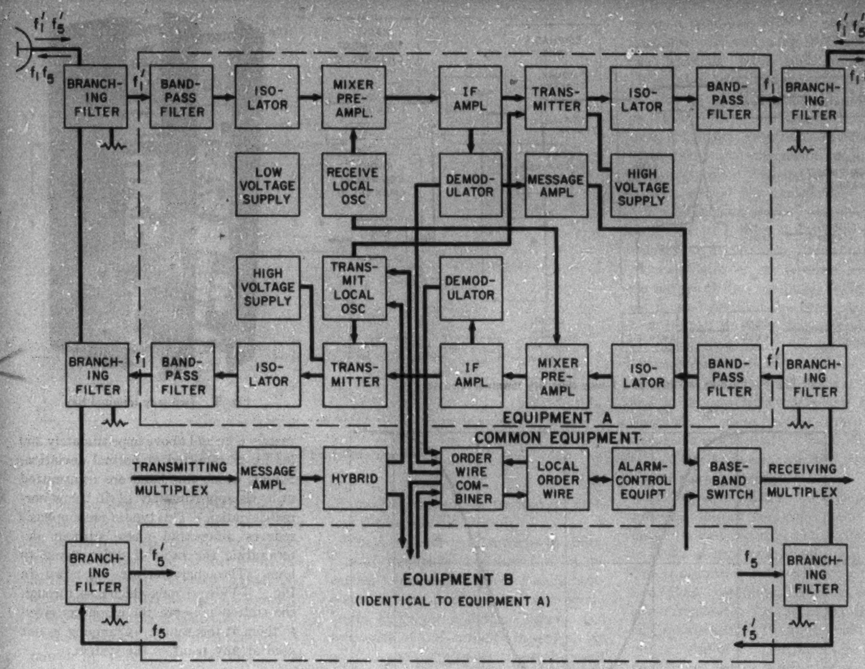

f_1' f_5'
f_1 f_5
BRANCHING FILTER
BAND-PASS FILTER
ISOLATOR
MIXER PRE-AMPL.
IF AMPL
TRANSMITTER
ISOLATOR
BAND-PASS FILTER
BRANCHING FILTER
LOW VOLTAGE SUPPLY
RECEIVE LOCAL OSC
DEMODULATOR
MESSAGE AMPL
HIGH VOLTAGE SUPPLY
HIGH VOLTAGE SUPPLY
TRANSMIT LOCAL OSC
DEMODULATOR
BRANCHING FILTER
BAND-PASS FILTER
ISOLATOR
TRANSMITTER
IF AMPL
MIXER PRE-AMPL
ISOLATOR
BAND-PASS FILTER
BRANCHING FILTER
EQUIPMENT A
COMMON EQUIPMENT
TRANSMITTING MULTIPLEX
MESSAGE AMPL
HYBRID
ORDER WIRE COMBINER
LOCAL ORDER WIRE
ALARM-CONTROL EQUIPT
BASE-BAND SWITCH
RECEIVING MULTIPLEX
BRANCHING FILTER
f_5'
f_5
EQUIPMENT B
(IDENTICAL TO EQUIPMENT A)
f_5
f_5'
BRANCHING FILTER

Fig. 5. Diversity drop repeater for message traffic

ing the message pilot unit. The voice channel test tone level at this point may be as low as −50 dbm. The pilot unit has two equal outputs, containing the input signal plus a pilot signal, each feeding a transmitter assembly. One transmitter assembly, including a message amplifier, modulator, and a transmitter unit, generates a radio carrier at, say, frequency f_1' and the other generates a carrier at frequency f_5'. The method by which these carriers are generated will be explained in the description of basic units, but it should be noted that the frequency deviation for a voice channel test tone is 200 kc rms. The nomenclature used for frequencies corresponds to that of the allocation plan in Fig. 1, channels 1 and 5 being chosen as consistent with the use of one polarization per feeder.

In the receive direction the branching filter directs incoming signals of frequencies f_1 and f_5 to the appropriate receiver assemblies. Each assembly contains a mixer, i-f amplifiers, a demodulator, and a message amplifier. The baseband switch selects the better of the signals it receives from the two message amplifiers and passes one of these, without attenuation, to the receiving multiplex equipment at a test tone level of −15 dbm. If pilot signals are present on both inputs, the pilot with the highest signal-to-noise ratio is selected but an input without a pilot is always rejected.

The ferrite isolators used in the transmitting and receiving r-f circuits are essential to the operation of this system. They remove the need for critical adjustments in the radio equipment, e.g., matching of traveling wave tubes, while maintaining a good match at the side presented to the branching network. The mechanical design of the radio equipment has been simplified by the use of isolators as will be explained.

A Drop Repeater Station

The block diagram for the associated diversity message dropping repeater is shown in Fig. 5. Incoming and outgoing radio frequencies are transmitted by branching filters, band pass filters, and isolators identical to those used in the terminal station. An incoming signal at f_1' reaches the mixer preamplifier typically at a level of −30 dbm. The signal is converted to intermediate frequency, amplified, and reconverted to a radio frequency of, say, f_1. In the 2-frequency system shown the receive local oscillator feeds two mixer preamplifier units. Similarly the transmit local oscillator feeds two transmitters. The local oscillators are crystal controlled providing excellent stability in the frequency translation process. Since the receive and transmit local oscillator frequencies are independent at both repeater and terminal stations a wide variation in the value of the translation frequency is possible. The standard translation is 213 mc. Other applications exist, however, where operation in a 400 mc band as prescribed by the CCIR is not desirable. These applications require frequency translations as low as 80 mc and as high as 300 mc. The use of a 2-frequency system (two r-f frequencies per 2-way radio channel) is made possible by the use of an antenna system with excellent front-to-back ratio.

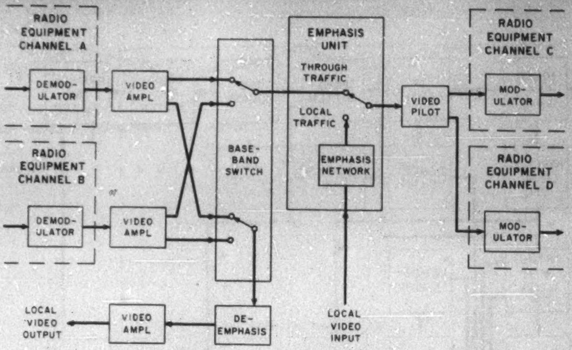


Fig. 6. Video facilities of a 1-way diversity terminal repeater

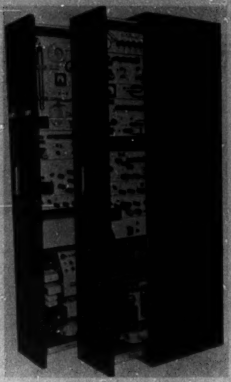


Fig. 8. Two-way terminal bay

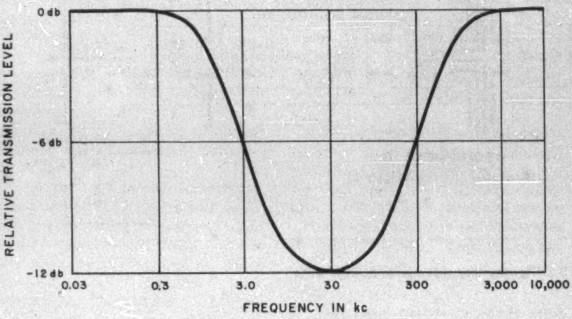


Fig. 7. Video emphasis characteristic

To provide for adding and dropping small amounts of traffic at repeater stations without the requirement for back-to-back terminal operation, a phase modulator is provided in the transmit local oscillator. When driven by a message amplifier this modulator is capable of adding to the outgoing carrier up to 60 voice channels in the 60 kc to 300 kc portion of the baseband spectrum. The received signals for these 60 channels are, of course, handled by the demodulator and receiving message amplifier. This low capacity drop traffic has the full advantage of diversity reception.

The service channel equipment for both terminal and repeater stations is designed so that one equipment assembly can serve the radio equipment for up to six 2-way radio channels. Service channel signals are distributed by the order wire combiner unit and are frequency modulated on the outgoing carrier by a second phase modulator in the transmit oscillator unit.[7]

Video Terminal Repeater

The organization of systems in which video traffic is transmitted is somewhat different from that illustrated in Figs. 4 and 5. The service channel signals, for example, are transmitted by a facility other than the video channel radio system. This removes the requirement for all demodulators and message amplifiers, at through repeater stations.

An illustration of the means by which video traffic is handled is shown in Fig. 6. This station is one in which diversified video traffic is received, used locally, and also retransmitted along the circuit. If desired, local video signals may be added. Other video station assemblies may be formed using the same equipment units. The emphasis network modifies the video signal such that signals below approximately 3 kc and above approximately 300 kc are transmitted at normal deviation. Signals inside this band are transmitted at levels approximately 12 db below normal deviation. This type of pre-emphasis reduces differential phase without deteriorating the ratio of video signal to hum. The characteristic is shown in Fig. 7. Video signals which pass through the station preserve the emphasis given to them at the source. Clamping is not used at any point in the system.

Description of Basic Units

Mechanical Design

The equipment uses cabinets which contain sliding racks. These racks are of standard width, holding 19-inch panels on which the basic units are built. A radio terminal, which is contained in two cabinets, is shown in Fig. 8. Four cabinets are required for a diversity terminal, excluding service channel supervisory equipment, and six cabinets are required for a diversity 2-way repeater including service channel and drop facilities.

A 2-way repeater is arranged so that one rack contains all the units which handle the through signal for one direction, i.e., a receiver and a transmitter. Between two such racks is a rack containing common equipment such as oscillators and service channel units.

Isolators and branching filters are mounted above the cabinets and connected to the pull-out racks by flexible coaxial cable. The branching filters are of waveguide construction but have coaxial outlets. The cabinet with two sets of branching filters can be installed in a room with only an 8-foot clearance from floor to ceiling.

Vol.
Pt. 1
1959

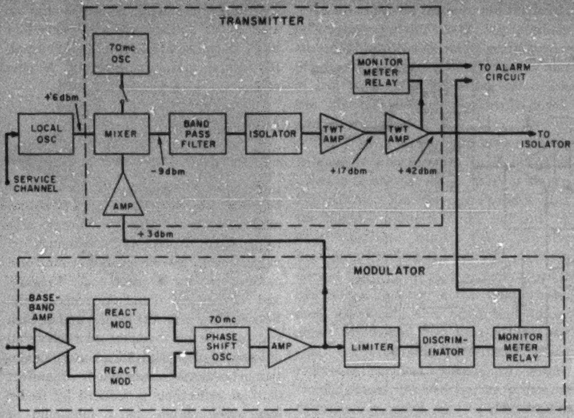


Fig. 9. Terminal transmitter assembly

The operation of the equipment may be monitored without the need for sliding an equipment rack out of the cabinet. For this purpose the cabinet front has meters indicating received signal and transmitted power as well as lights to indicate the status of controls.

More detailed information may be obtained by pulling out the rack. The equipment remains in an operating condition; all units being shielded for the protection of the maintenance personnel. In particular each rack and unit is provided with facilities for the measurement of tube currents and, by depression of the filament voltage, the emission capabilities of each tube may be ascertained. This is a facility for preventative maintenance which is of great value and particularly effective if records of previous measurements are kept. This facility will reduce operating tube failures and unnecessary replacement of tubes.

The use of individual unit rear covers allows circuits to be checked and repairs to be made in one unit mounted in the rack without accidental contact with other units. Also the removal of a unit is facilitated by the use of reliable plugs for signal and power connections.

The cabinet construction saves floor space while providing simultaneous accessibility to both sides of each unit. Cabinets may be mounted side by side and placed against a wall or may be placed back to back.

Terminal Transmitter

The radio transmitter block diagram is shown in Fig. 9. It consists of 3 units: the modulator unit, the transmitter unit, and the local oscillator unit.

The modulator unit performs the function of frequency modulating a 70 mc carrier. Included in the unit are a baseband amplifier, push-pull reactance modulators, and a phase shift oscillator. The frequency of the oscillator is very stable so automatic frequency control is not necessary. The stability is due to the balancing effect of the push-pull modulators. A 70-mc discriminator and meter are used to monitor the center frequency of the oscillator.

The slope of the conversion characteristic of a typical modulator varies less than 1% over a frequency deviation range of ±4 mc; the overload points being at ±12 mc.

The transmitter unit contains an i-f amplifier, a low level balanced crystal mixer, a band-pass filter for selection of the proper sideband, a ferrite isolator, and two traveling wave tube amplifiers.[8] These tubes have been specially designed for this equipment and are completely packaged. They are designed for extremely long life; requiring no mechanical adjustments and only minor voltage adjustments, upon installation.[9] Fig. 10 shows a view of this transmitter.

The mixer is an up-converter which obtains excitation from the local oscillator unit and signal from the 70-mc output of the modulator or, at repeater stations, the output of the i-f amplifier. The isolator is used between the band-pass filter and the first traveling wave amplifier to provide a good match for the filter. A 70-mc crystal-controlled oscillator, using transistors, is included in the transmitter and is automatically connected to the mixer should the i-f signal from the receiver in a repeater fail. This permits the transmitter to operate for the transmission of alarm signals.

Each local oscillator unit will provide sufficient output for either two transmitters or two receivers. Each contains a crystal oscillator and a multiplier chain. The transmitter local oscillator contains two phase modulator circuits, one at audio frequencies for the service channel and one for adding traffic channels at repeater stations. The audio-frequency modulator circuit consists of a ferrite line of variable-delay and the traffic channel modulator uses a conventional tube circuit. De-emphasis networks preceding each phase modulator produce the desired frequency modulation characteristic.

Terminal Receiver

The receiver is composed of four units: mixer-preamplifier unit, i-f unit, demodulator unit, and a baseband amplifier

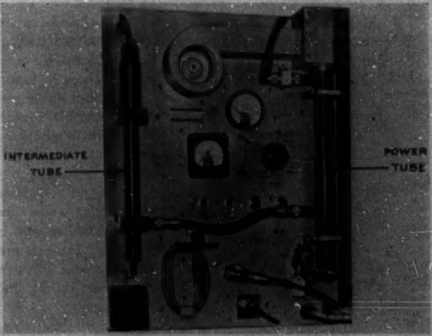


Fig. 10. Transmitter panel showing traveling wave tubes

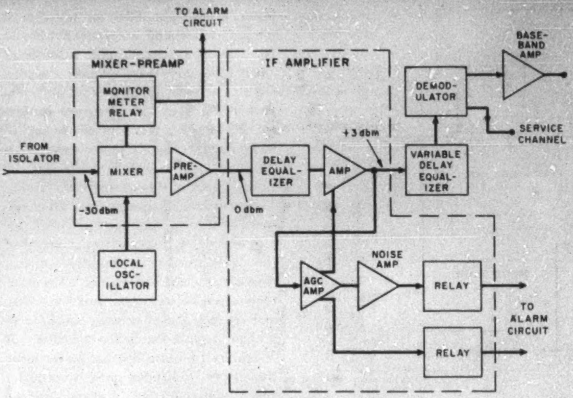


Fig. 11. Terminal receiver assembly

unit. The complete receiver block diagram is shown in Fig. 11.

The mixer-preamplifier unit contains an unbalanced crystal mixer and four stages of amplification at 70 mc. The first two stages use grounded grid circuits. A noise figure of between 8 and 10 db is obtained.

The i-f unit contains a fixed delay equalizer and seven stages of amplification using conventional double tuned interstage transformers.[10] The Q ratio of these transformers is designed to provide maximum band-pass stability with AGC action and tube changes. Additional circuits in this unit provide automatic gain control, and give alarm information should the receiver gain drop or the received signal fail. The signal failure circuit also operates a noise squelch and, at repeaters, turns on the 70 mc oscillator in the transmitter. Two outputs are provided, one for connection to the demodulator and one for connection to the transmitter at repeater stations.

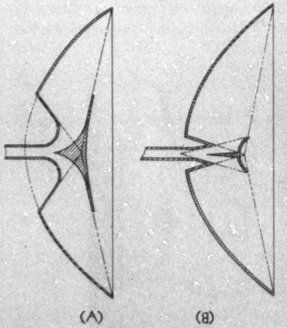


Fig. 12. Evolution of circular horn-paraboloid

A—Original
B—Modified

The demodulator unit includes a limiter driver tube, two limiter stages, a harmonic filter, and parallel discriminator drivers. The discriminator is a symmetrical arrangement of two tuned circuits which provides excellent balance and at the same time a single-ended output. Following the discriminator is a cathode follower for the base-band output and a 2-stage audio amplifier for the service channel.

The over-all gain of the receiver is sufficient to provide adequate limiting on signal strengths down to FM threshold.

Either one of two baseband amplifier units may be used. One is suitable for monochrome or color TV and the other, a simplified version, is suitable for message traffic only. Both amplifiers use negative feedback over three stages.

The baseband switching unit contains circuits for the selective amplification and detection of the radio pilot frequency present at each input. The switch is a transistor device driven by a trigger circuit and has a switching time of approximately 10 microseconds.[11]

Two types of d-c power supply are used: a low-voltage supply for the receiver circuits and a high-voltage supply for the traveling wave tube amplifiers. Silicon rectifiers are used in all critical circuits. Circuit breakers are used to protect the transmitter tubes against excessive helix current during initial adjustments of voltages.

All inter-unit connections at baseband frequencies and at 70 mc use 75-ohm coaxial cables, and all units are designed for high return loss at each input and output. R-f connections are made by 50-ohm cables.

Antenna System

Performance Requirements

Both the antenna and the feeder must be efficient in order that a good signal-to-noise ratio be obtained with reasonable transmitter power and antennas of modest size. Economic limits for these parameters were assumed to be 15 watts and 10 feet in diameter, respectively. Reflections within the feeder, and at each end, must be low to reduce color television differential phase and reduce intermodulation between voice channels.[12] The effect of these reflections is reduced by feeder loss which should be a minimum of about 0.4 db per 100 feet assuming an antenna of VSWR (voltage standing wave ratio) 1.06 maximum. The maximum permissible loss for good signal-to-noise ratio with an antenna of high efficiency is about 1.5 db per 100 feet.

The directional properties of the antenna must be controlled carefully. Very little energy must be radiated in the back direction as this energy interferes with transmission in the opposite direction.[13] Radiation at small angles off the main beam must be low to avoid "overshoot," a phenomenon by which a signal may bypass intermediate repeaters.

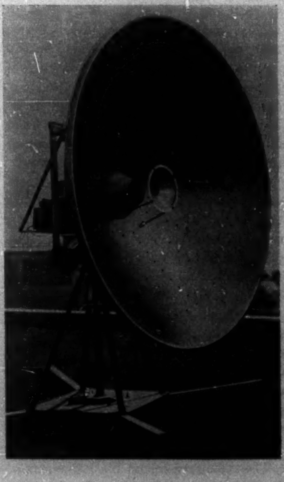


Fig. 13. Antenna

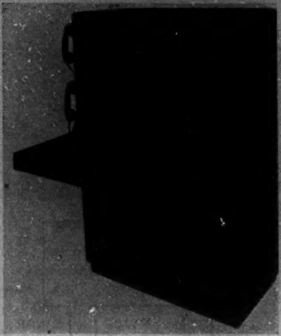


Fig. 14. Console containing all supervisory station service channel facilities

Also, the cross-polarized energy radiated in the direction of the main beam must be low to avoid interference with adjacent r-f channels of the CCIR frequency plan. For one antenna to serve three transmitters and three receivers on a common polarization it must have a minimum bandwidth of 400 mc.

Feeder Design

For 2,000-mc operation there are two types of feeder which may be used: "semiflexible" coaxial cable and waveguide.[14] In coaxial cable, freedom from the transmission of spurious modes sets the maximum useable standard diameter as $1^5/_8$ inches. This size has an attenuation of the order of 1.4 db per 100 feet but the magnitude of internal reflections suggest that its use be limited to systems, of 240-voice channel capacity or less. It is, however, suitable for monochrome television transmission.

The *MM-600* uses rectangular waveguide of outside dimensions 4.30 by 2.15 inches. Aluminum waveguide is used for ease of handling and had a measured attenuation of only 0.5 db per 100 feet. The feeder and the antenna are pressurized to a maximum of 2 psi (pounds per square inch).

Antenna Design

Conventional antennas, using a shallow paraboloid illuminated by a waveguide or coaxial horn, do not have the required VSWR over a ±15% band and have excessive back radiation. The horn-reflector antenna, which was designed specifically for radio relay use at frequencies of 4,000 mc and above, does have all the electrical properties desired but is large and heavy.[15–17]

An attempt was made to use the horn-reflector concept and modify it to produce an antenna with acceptable performance of larger aperture area, 80 square feet, and of less bulk. The procedure by which the *MM-600* antenna was evolved will now be explained.

The frontal area of the horn-reflector antenna is a sector of a circle but it is possible to use the whole circle as shown Fig. 12(A). For the same useful area the height is reduced to one half and the depth to one third. This "circular horn-paraboloid" is fed from the rear by a curved horn of coaxial cross section which blocks one quarter of the total frontal area. Reducing this large feed to increase the useful area also reduces the side lobes of the radiation pattern. Using reflecting surfaces instead of bends, simple shapes, derived by the principles of geometric optics, may be employed in the feed. The antenna, as shown in Fig. 12(B), then resembles a Cassegrain telescope. Further modifications from these shapes were found to be necessary to obtain good impedance match. For example the central pin was virtually eliminated. Fig. 13 shows the end product: a deep paraboloid $(f/d=0.2)$ of 10 feet in diameter. A window has been added to the mouth of the circular waveguide horn and behind the dish there is a transition from circular to rectangular waveguide.

The gain varies from 32.6 db at 1,700 mc to 36.5 db at 2,300 mc, the corresponding beam widths varying from 4 to 2.8 degrees. A VSWR of 1.03 at 2,000 mc is obtained.

The Service Channel

The maintenance of a long-haul communications system is a major operating expense. A comprehensive service channel system will decrease the cost of maintenance while increasing system reliability. Full facilities for unattended system operation may be provided by the service channel equipment to be described. The equipment arrangement is flexible and expandable to meet specific requirements.

The term "service channel" should be explained. It is used to denote all communication, data transmission and processing, display and control facilities provided for the maintenance of a radio relay system. Such a definition includes the communications means, which may be via a separate radio link, by wire lines, or over the relay system itself.

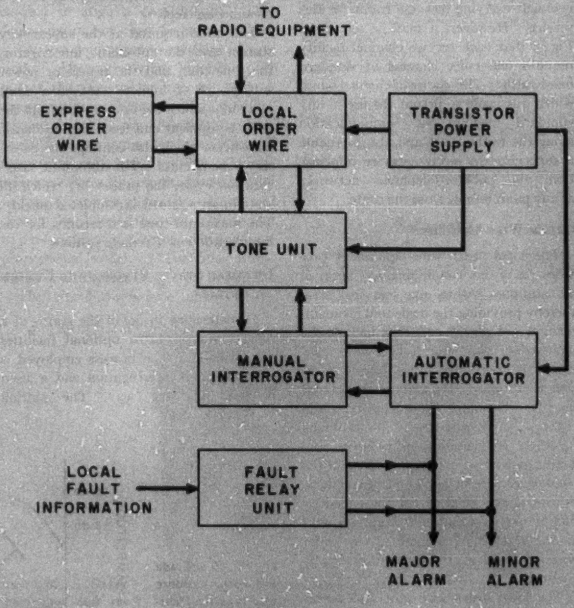


Fig. 15. Supervisory station

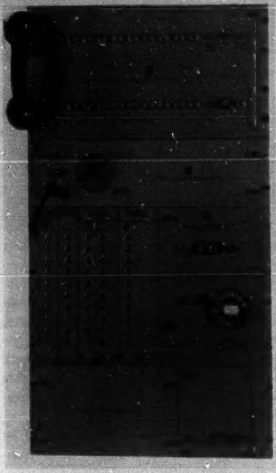


Fig. 16. Service channel equipment at a remote station

The design of the *MM-600* service channel does not include a special radio channel but uses one or more of the radio channels carrying message traffic for this service. However, it may be seen from Fig. 2 that each service channel facility requires one voice channel of standard bandwidth. This choice of channel bandwidth was made to permit the use of any standard medium for communication along the traffic route and also to enable a supervisor to receive service information via public telephone networks at any point remote from the route.

ORDER WIRE EQUIPMENT

The local order wire equipment provides party-line communication between all stations within one service area, thereby providing the basic tool for maintenance. A service area may include the main route plus spur links but is commonly 1 supervisory station plus 3 to 15 remote stations. The maintenance center may be at a point remote from the radio route; 4-way 4-wire hybrids are provided for these connections. Also 2-wire and 4-wire extensions are incorporated in the local order wire unit. Loudspeaker hailing and tone signalling may be added by plug-in units. Combining units are used if, for reasons of reliability, the order wire is carried over more than one radio channel.

The local order wire unit also serves as a distribution unit for the signals of other facilities including the express order wire which is intended for communication between maintenance centers. The express unit is similar to the local unit with the addition of translation circuits for the 8.3- to 11.7-kc band.

MINIMUM ALARM FACILITIES

It is a primary requirement that a supervisory station receive accurate information on the status of remote equipment. The degree of detail required will vary from system to system but a convenient design minimum must be set. The automatic reporting of four status conditions per remote station, which station is positively identified, was considered to be the minimum requirement. A "fail safe" design was considered mandatory so that no fault could occur without identification. These objectives were achieved by the use of the following basic principles:

1. Interrogation by the supervisory station.
2. Allocation of a separate tone channel by which each remote station receives the interrogation pulse and gives a coded reply.

The four status conditions may be called: no-alarm, major-alarm, minor alarm, and alarm system-alarm; the latter being the interpretation given to a meaningless reply.

The units required at the supervisory station are: the automatic interrogator, the tone unit, and the transistor power supply. At a remote station a fault relay unit stores the information from the radio equipment and feeds the automatic responder, which also contains the necessary tone circuits. The automatic interrogation cycle, the process by which all stations are scanned, is completed quickly. The maximum time is 6 seconds for the interrogation of 15 remote stations.

DETAILED STATUS REPORTS AND REMOTE CONTROL

To determine in detail the status of a remote station extra optional facilities are available. The process employed is called manual interrogation and is used in the following way. The station selected is withdrawn from the automatic interrogation cycle which continues to survey the other stations. A statement from the selected station is requested and by an electronic stepping process a maximum of 20 conditions is automatically reported and displayed visually at the supervisory point.

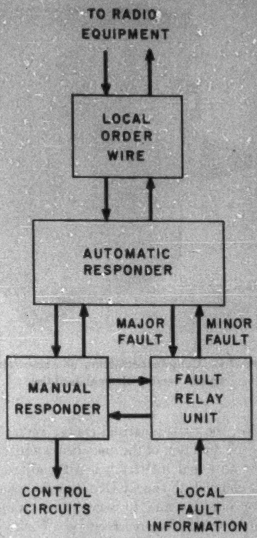


Fig. 17. Remote station

Remote control facilities are available as an extension of the manual interrogation process. After a control function is selected, a report from the remote station, confirming that the proper function has been selected, may be received and displayed. Then by the transmission of a special tone the control is executed. Confirmation of an order completed is of course obtainable at any time by manual interrogation.

The manual interrogator and responder units accept plug-in assemblies to provide continuous expansion of facilities for 1 to 15 stations each with 10 or 20 reports and 0 or 10 controls. Any of the num-

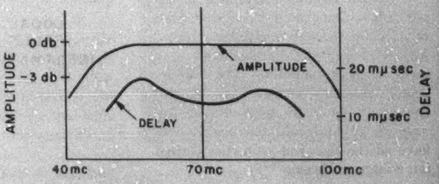


Fig. 18. Amplitude and delay response of a typical I-f network

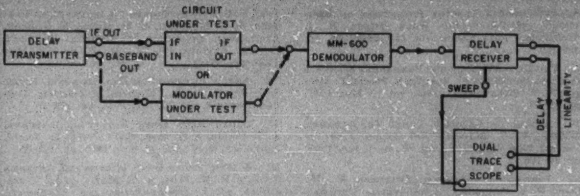


Fig. 19. Simultaneous measurement of delay and linearity

bers given above can be increased by the use of additional main units. A supervisory station equipped with maximum standard facilities is shown in Figs. 14 and 15. The corresponding remote station equipment, which can be incorporated in the radio bays, is shown in Figs. 16 and 17.

The service channel system, complete within itself, has many applications apart from the maintenance of radio relay systems. Extreme reliability is obtained by the application of sound basic operating principles to equipment which incorporates special long-life components including transistors and cold cathode tubes. No mechanical components are used for repetitive switching functions. Instead techniques taken from computer technology are employed.

The concept and planning of this system was a major achievement for which Mr. G. Birman should receive special credit.

Test Equipment

Test facilities are incorporated in the design of the radio equipment itself and some of these were described in the section on basic units. Special test points have also been included; for example, the feeder system includes directional couplers for the measurement of VSWR and station gain flatness. Sweep generators and amplitudes detectors for r-f and i-f use are essential items and other special equipments are needed.[18] Of particular interest and great importance is the group delay measuring equipment.

All r-f and i-f band-pass filters used in the radio system are designed to have maximally flat amplitude response and therefore do not have a phase characteristic linear with frequency. Group delay, the slope of this phase characteristic, is not constant and must therefore be measured and equalized. The amplitude and delay response of a typical i-f network is shown in Fig. 18. Considerable accuracy of equalization is required and delay measurements must be made on an operating system to ensure this.

The linearity of the transfer characteristic of a system must also be measured. In particular, modulators and demodulators require alignment by so-called derivative techniques. The delay equipment described incorporates all the circuits necessary for linearity measurements to be made simultaneously with delay measurements.

The delay and linearity measurement equipment is built in the form of two independent units. The transmitter unit supplies effectively distortionless signals and the receiver unit analyses the distortion introduced by intermediate equipment under test. This arrangement is illustrated in Fig. 19, which shows several important features. The transmitter unit has an i-f output, modulated by a 200-kc tone and swept across the i-f band, which is used to align a demodulator. This demodulator can then be used as part of the test equipment to measure i-f networks. Alternatively the baseband output of the transmitter may be used to measure the performance of a modulator and demodulator. For checking purposes the direct connection of receiver and transmitter units is also possible. The receiver measures the amplitude of the 200-kc tone to provide a linearity output, and detects the phase of the tone, with respect to a reference signal, to provide a voltage proportional to delay. The delay transmitter and receiver are shown in Fig. 20.

The signal-of-reference phase is derived from the input signal itself so that delay measurements can be made over 1-way multi-hop circuits. A sweep voltage for the oscilloscope is also taken from the input signal. The equipment is capable of measuring delay variations of 1 millimicrosecond and changes of slope of transfer characteristics to less than 1/2 of 1%.

Summary

The careful attention paid during the design period to user requirements and the completeness of the design make the *MM-600* system suitable for the varied needs of commercial and military high-capacity communication networks. Systems were under test in the laboratory in the beginning of 1958 and a 2-hop experimental field circuit has been operated since September 1958.

Appendix. Technical Specifications

Frequency band	1,700 to 2,300 mc
Length of system	4,000 miles
Relay station spacing	30 miles (average)
Type of modulation	FM
Maximum deviation	±4 mc peak
Type of repeater	heterodyne with 70 mc intermediate frequency
Transmitter output power	15 watt minimum (+42 dbm)
Antenna gain	34 db
Receiver bandwidth	30 mc
Receiver threshold	−79 dbm (10 db noise factor)
Fading margin (30-mile path)	49 db
Capacity of each r-f channel	600 voice channels plus service channel, or one television signal
Add channel provision at repeaters	60 voice channels, 60 to 300 kc

References

1. The Planning of National Telecommunications Networks, C. A. Parry. *AIEE CP 58-236* (available on request).

2. Some Results with Frequency Diversity in

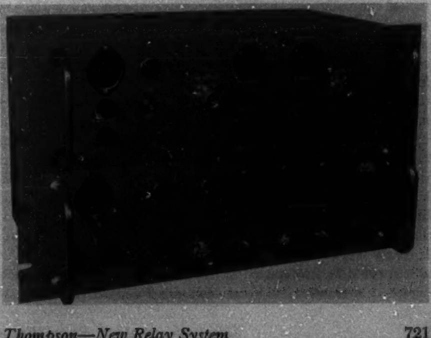


Fig. 20. Delay transmitter and receiver

A Microwave Radio System, **F. H. Willis.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, Mar. 1957, pp. 63–67.

3. Tele-Control of Radio Link FHT 4003, **M. Rastello.** *Onde Electrique*, Paris, France, vol. 37, no. 368, Nov. 1957, pp. 1018–21.

4. Distortion in Multichannel Frequency-Modulation Relay Systems, **L. E. Thompson.** *RCA Review*, Princeton, N. J., vol. 11, no. 4, Dec. 1950, pp. 435–64.

5. Non-linear Distortion in Transmission Systems, **R. A. Brockbank, C. A. A. Wass.** *Journal*, Institution of Electrical Engineers, London, England, vol. 92, pt. III, Mar. 1945, pp. 45–56.

6. Automatic Protection Switching for TD-2 Radio System, **G. H. Klemm.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, Nov. 1956, pp. 520–27.

7. The Life and Reliability of Valves, **K. Rodenhuis, H. Santing, H. J. M. Van Tol.** *Phillips Technical Review*, Eindhoven, Holland, vol. 18, no. 7, Dec. 1956, pp. 181–92.

8. Effect of Collector Potentials on Efficiency of Travelling Wave Tubes, **H. J. Wolkstein.** *RCA Review*, vol. 19, June 1958, pp. 259–82.

9. The Design of Periodic Permanent Magnets for the Focussing of Electron Beams, **F. Sterger, W. W. Siekanowicz.** *Ibid.*, vol. 18, Mar. 1957, pp. 39–59.

10. A Note on Parallel-Tuned Transformer Design, **V. C. Rideout.** *Bell System Technical Journal*, New York, N. Y., vol. 27, no. 1, Jan. 1948, pp. 96–108.

11. Transistor Bilateral Switches, **W. M. Cook, P. L. Bargellini.** *Semi-Conductor Products*, vol. 1, no. 4, July-Aug. 1958; no. 5, Sept.-Oct. 1958.

12. Interchannel Interference in FM and PM Systems Under Noise Loading Conditions, **W. R. Bennett, H. E. Curtis, S. O. Rice.** *Bell System Technical Journal*, vol. 34, May 1955, pp. 601–36.

13. Distortion in Frequency-Division-Multiplex FM Systems Due to an Interfering Carrier, **R. G. Medhurst, E. M. Hicks, W. Grossett.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 105, pt. B, no. 21, May 1958, pp. 282–92.

14. Aerial Feeders for Multichannel Links, **L. Lewin, J. Payne.** *Electronic Engineering*, London, England, vol. 30, July 1958, pp. 414–19.

15. Broad Band Horn Reflector Antenna, **A. T. Corbin, A. S. May.** *Bell Laboratories Record*, New York, N. Y., vol. 33, Nov. 1955, pp. 401–05.

16. A New Broad-Band Microwave Antenna System, **R. W. Friis, A. S. May.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, Mar. 1958, pp. 97–100.

17. Horn-Paraboloid Aerials for Wide-Band Radio-Link Systems, **H. Laub, W. Stohr.** *Frequenz*, Berlin, Germany, vol. 10, no. 2, Feb. 1956, pp. 33–44.

18. Maintenance Equipment for the GDH 103 Radio Link, **M. Colas.** *Onde Electrique*, vol. 38, no. 374, May 1958, pp. 369–77.

A High-Volume High-Speed Weather Information Distribution System

E. E. SCHWENZFEGER
ASSOCIATE MEMBER AIEE

THE PLANNERS of aircraft operations, both commercial and military, have always required up-to-date information regarding predicted weather along flight paths and at destinations. The need for more timely weather information has been increasing for some years. On the commercial side the need has grown because of the increasing density of air traffic and the economic drive for more efficient aircraft operations. For the military side, apart from the constant drive for greater efficiency, the need has also been growing because of the increasing complexity of operations.

The advent of jet aircraft has made the need urgent for more timely weather information. This urgency stems in part from the fact that a jet consumes such a large amount of fuel at low altitudes that a missed approach becomes a very expensive maneuver. The quality of weather information generally available today obliges the commercial operator to plan for the contingency of finding the end of the runway shut in after descending to an uneconomically low altitude. This planning calls for carrying sufficient extra fuel to permit regaining an economical altitude and reaching an alternate field or holding while waiting for the ceiling to lift. The extra fuel carried translates directly into a reduction in pay load. If timely weather information were available it could be transmitted to the plane before it descended to lower altitudes. In military operations good predictions regarding end-of-runway weather are especially important because jets returning from extended missions may have too little fuel left to permit any change of plan after reaching a low altitude. It will, of course, be appreciated that the foregoing is only one facet of a many-sided problem.

Weather system planners in the CAA (Civil Aeronautics Administration), U. S. Weather Bureau, and the Air Force Cambridge Research Center have been working on what may be called a 2-pronged attack on the problem. One prong calls for the development of sophisticated hardware for the rapid acquisition and dissemination of weather data. The other prong calls for the development of numerical methods employing computers for making forecasts. There are several numerical approaches which might be taken. One is a statistical method which requires reiterated solutions of a mathematical model utilizing constantly acquired data; another requires solutions of a thermodynamic model of the atmosphere and a third requires the solution of equations of motion. The equations for the third approach, which were furnished by D. R. Harris, Chief of Weather Bureau's Aero-Meteorological Investigations, are shown in the Appendix. These are offered only to give the flavor of the computation problem. These equations must be solved for each 100-mile grid point spread over the continent and adjacent waters. It is understood that a 24-hour forecast requires 144 solutions spaced 10 minutes apart for each grid point.

Because observation and forecast data are very perishable, rapid and efficient communication is a basic requirement in instrumenting the 2-pronged attack.

The communication system presently beging used to disseminate weather data is a conventional 75-word-per minute teletypewriter system. The system is composed of three services: Service A is for domestic aviation, Service C is for general weather service, and Service O is for international aviation. The demands made upon this system have for some years far exceeded its capabilities. A characteristic of the present system is that users at many receiving stations are obliged to receive a great deal of unwanted data in order to get what they need. In extreme cases the unwanted data run as high as 80% of the total received. Also, the system is inflexible in that when a user wishes to alter the program of received data he must put a request through Washington for the change; this requires several weeks to place in effect.

The Air Navigation Development Board in conjunction with the CAA asked the Bell Telephone Laboratories to make an engineering study and propose a system which would satisfy present-day and future requirements. This work was done under Signal Corp Contract *DA36-039 SC-64567* and was reported on under the title "Air Traffic Control Communications, Final Report, Task D Part 2—

Paper 59-812, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing April 20, 1959.

E. E. Schwenzfeger is with the Bell Telephone Laboratories, Inc., New York, N. Y.

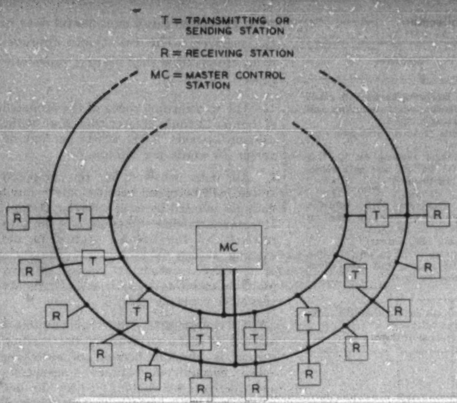

T = TRANSMITTING OR SENDING STATION
R = RECEIVING STATION
MC = MASTER CONTROL STATION
MC
T
R

Fig. 1. Proposed system

High-Volume Weather Information Communication System." Permission to discuss the proposed system in this paper has been granted by the Aviation Weather Division of the Federal Aviation Agency which now has the responsibility for this work. The military services, and particularly the Cambridge Air Research Center of the Air Force, took an active part in formulating these requirements, some of which are stated below to give the flavor of the communications problem.

1. The system shall be capable of accomplishing the collection and distribution of meteorological and NOTAM (Notice-to-Airmen) data on a national basis with a means of automatic interconnection with the weather and NOTAM system of contiguous countries and conventional teletypewriter systems within the United States.

2. The system shall be capable of accommodating the existing load on the Services A, C, and O teletypewriter networks, the USAF (United States Air Force) national teletypewriter networks and the special weather teletypewriter networks of the U. S. Weather Bureau. It is estimated that the load will increase to 1,500,000 groups per day in 1965 and to 2,050,000 groups per day in 1975. A group is a 5-character word plus a space.

3. The maximum message unit transit time of programmed data from origin to destination anywhere in the United States must be less than 10 minutes.

4. Provision shall be made for the receivers to determine their needs and be able to select and program data by types from a minimum of 500 selections. The ability to alter the selection on an hour-to-hour basis is a necessary feature.

5. The end product at receiving stations must be printed page copy with provision for feeding data directly, without human intervention, into numerical weather computing machines.

6. Consideration should be given to the possibility that the operating personnel at both receiving and transmitting stations will not always be experienced in the operation or adjustment of communication equipment.

General Description of the Proposed System

The system proposed to meet the above requirements is a high-speed distribution system, shown in Fig. 1, and is in effect based on the old telephone party line concept in which any or all subscribers could listen in on any or all conversations if they wished. It is a party line arranged for transmitting data and messages at the rate of 1,000 words per minute. There may be as many as 1,000 stations which transmit data and perhaps as many as 5,000 which receive them.

However, the proposed system is unlike the old telephone party line in that the transmitting station gets on the line only at the invitation of a master control station which is programmed by operating personnel to call in sending stations as required. Provision is made for the master control to give any sending station emergency access to the line. Stations, before sending data, identify themselves and the data they are going to send. This also departs somewhat from the old party line principle. Each receiving station is programmed, that is to say, instructed by its own local operating personnel to monitor the line and to attach its printer to the line automatically to receive desired data from transmitting points in which it is interested. The program at receiving stations may be arranged to select any number of transmissions from the line. This program may also be readily altered by nontechnical operating personnel to meet changing data requirements.

The telephone facilities which make up the long distance communication network of this country are arranged to connect towns and cities. Therefore, the circuit layout of the proposed system must consist of a series of lines fanning out arterially from interconnected hubs scattered about the country. The number and locations of the hubs present a problem of considerable magnitude to the circuit layout engineer because of the many economic and technical considerations involved.

Fig. 2 shows in schematic fashion the principles involved in laying out the system. It does not represent the actual layout which might be used.

The manner in which data are distributed over the network may be seen by considering the route by which data leave

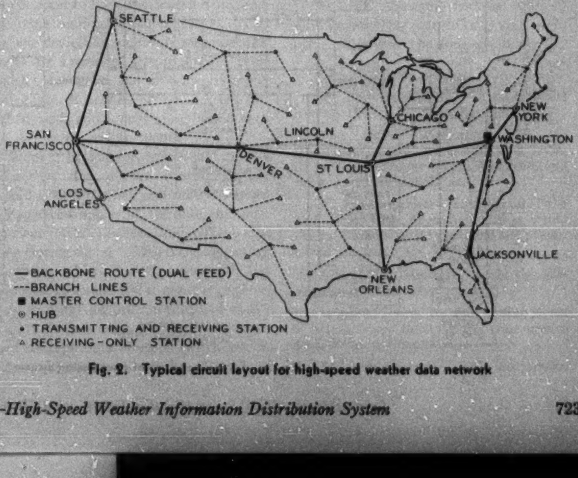


Fig. 2. Typical circuit layout for high-speed weather data network

Table I. Mode of System Operation: Scheduled Collection

Master Control (Mc)	Station Transmitter	Receiver
Sends SA>SA>SA>SA date time <▬▷◁	All transmitters with tape data selector switch on SA and having SA cards cock to transmit	1. All receivers having SA cards trigger, open gate to line, and start on first SA> 2. Prints "SA>SA>SA date time" 3. Carriage returns on <, line shifts on ▬, and gate closes to line on ▷◁
1. Sends SADCA 2. Checks for receipt of DCA<▬▷◁ 3. Times 0.1 sec. for DCA 4. Times 1 sec. on quiet line for <▬▷◁ 5. If timing is not satisfactory, sends <▬▷◁	1. DCA transmitter cuts in and sends "DCA (text)" <▬▷◁ 2. Receipt of ▷◁ (its own ▷◁ or MC's) closes gate to line	1. All receivers with SADCA cards open gate to line on receipt of SADCA 2. Prints "DCA (text)" 3. Carriage returns on <, line shifts on ▬, and gate closes to line on ▷◁
Sends SAXXX, etc. At end of program, sends EP> EP▬▬▬	XXX transmits, etc.	All receivers having SAXXX, etc. 1. Prints "EP▬▬▬" 2. Printers shut down

Symbols:
> = space
< = carriage return
▬ line feed
▷◁ = gate closing signal
EP>EP▬▬▬} end of program signal

a particular transmitting point, say Lincoln, Neb. The data flow in several directions from Lincoln to all points on its branch line. One of these points is Denver, Colo., which is a hub point on the backbone route. The data fan out to the rest of the country from Denver.

The proposed system is efficient in the use of the line in that only 5% of the time required for an average data transmission is utilized for control and selection functions. Also, in the proposed system, unlike the existing system, data sent into the line from any transmitting station are immediately available to any receiving station that wants them.

Comparative traffic-handling capabilities may be judged from the fact that the combined maximum hourly load of Services A, C, and O, which constitute the present system, could be handled in less than 15 minutes by the proposed system. Also, because it confers full flexibility in choice of data received, the proposed system delivers 20 times more useful data to the receiver. The twentyfold improvement grows out of the following considerations:

1. The transmission speed of the proposed system is 13 times greater than that of the existing system: 1,000 words per minute versus 75 words per minute.

2. All data received in the proposed system will be useful because only wanted data are selected by each receiving station. It is estimated that at least 25% of the data received by the average station in the existing system is unwanted. The ratio of effective transmission speeds is thus probably closer to 1,000 words per minute versus 56 words per minute.

3. A considerable but undetermined amount of data received in the existing system is obsolete because of switching delays encountered between transmitters and receivers. Switching delays do not occur in the proposed system because there are no centralized switching points to introduce delays. Thus the transit time for data from origin to destination is less than a second, which is many times better than the requirement which calls for a transit time of less than 10 minutes.

Operation of the Proposed System

Before describing the system components in detail it might be useful to consider first the operation of the proposed system.

All transmitting stations on the line

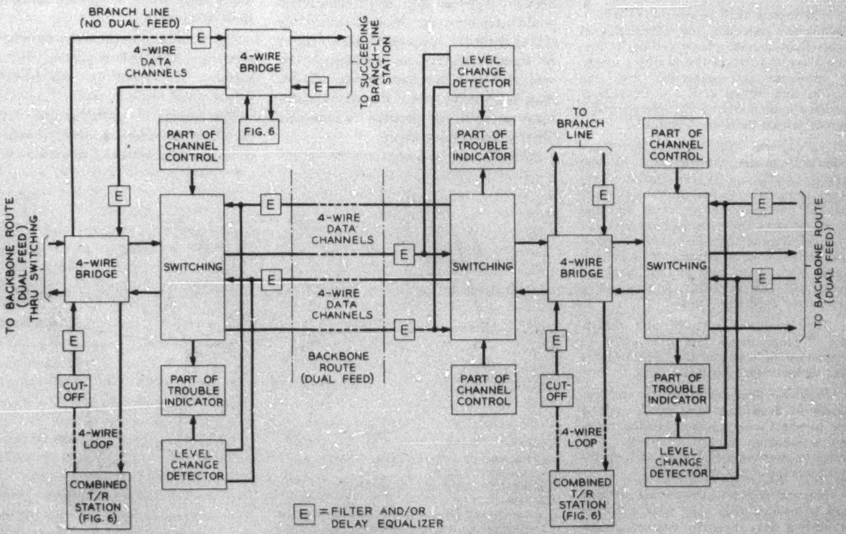

BRANCH LINE (NO DUAL FEED)
4-WIRE DATA CHANNELS
4-WIRE BRIDGE
TO SUCCEEDING BRANCH-LINE STATION
FIG. 6
PART OF CHANNEL CONTROL
LEVEL CHANGE DETECTOR
PART OF TROUBLE INDICATOR
TO BRANCH LINE
PART OF CHANNEL CONTROL
E
TO BACKBONE ROUTE (DUAL FEED) THRU SWITCHING
4-WIRE BRIDGE
SWITCHING
4-WIRE DATA CHANNELS
SWITCHING
4-WIRE BRIDGE
SWITCHING
TO BACKBONE ROUTE (DUAL FEED)
4-WIRE DATA CHANNELS
BACKBONE ROUTE (DUAL FEED)
PART OF TROUBLE INDICATOR
PART OF CHANNEL CONTROL
PART OF TROUBLE INDICATOR
CUT-OFF
CUT-OFF
LEVEL CHANGE DETECTOR
LEVEL CHANGE DETECTOR
4-WIRE LOOP
4-WIRE LOOP
COMBINED T/R STATION (FIG. 6)
COMBINED T/R STATION (FIG. 6)
E = FILTER AND/OR DELAY EQUALIZER

Fig. 3. Typical segment of transmission system

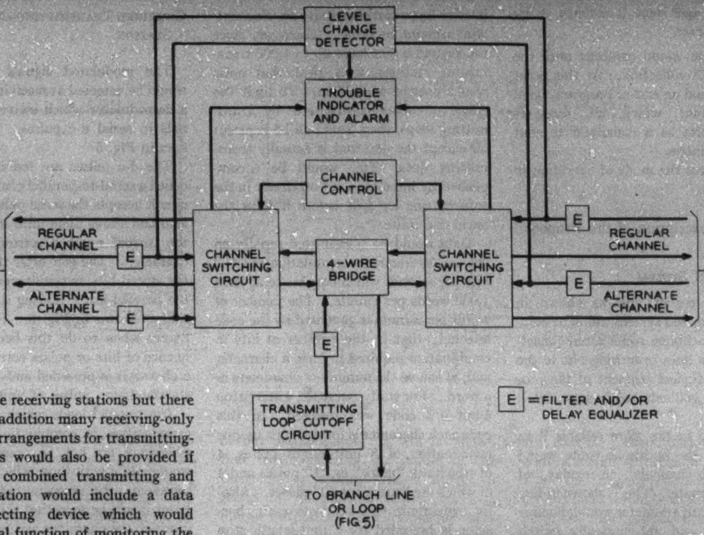

LEVEL CHANGE DETECTOR
TROUBLE INDICATOR AND ALARM
CHANNEL CONTROL
BACKBONE ROUTE (DUAL FEED)
REGULAR CHANNEL
ALTERNATE CHANNEL
E
CHANNEL SWITCHING CIRCUIT
4-WIRE BRIDGE
CHANNEL SWITCHING CIRCUIT
REGULAR CHANNEL
ALTERNATE CHANNEL
BACKBONE ROUTE (DUAL FEED)
TRANSMITTING LOOP CUTOFF CIRCUIT
TO BRANCH LINE OR LOOP (FIG 5)
E = FILTER AND/OR DELAY EQUALIZER

Fig. 4. Channel equipment on backbone route

would also be receiving stations but there would be in addition many receiving-only stations. Arrangements for transmitting-only stations would also be provided if desired. A combined transmitting and receiving station would include a data program-selecting device which would have the dual function of monitoring the line for both control codes and data program code. A data program code is a unique combination of characters which prefaces the data sent by each transmitter; it tells the data program selectors what type of data is being sent by which transmitter. Control codes are other character combinations used to start transmitters, indicate end of transmission, indicate end of program, etc. The selecting device would translate these codes and feed information to control logic circuitry. The logic circuitry would connect and disconnect the printer to and from the line to receive only the desired data which corresponds to decoded data program codes. It would also start the tape transmitter under command of a master control station and perform other functions as directed by decoded control codes.

There would be one master control station on the entire line. Its function would be to control the operation of the system. It would be programmed or instructed to start transmitters automatically and command them to transmit particular data at given times. The control station would also detect and correct certain malfunctions and call in operating personnel to deal with troubles it is unable to handle.

On the present Teletype weather network there are 8 scheduled and 27 unscheduled data collections (transmissions) which are made on a periodic basis. The collections are identified by designators such as:

SA for hourly aviation weather reports
SI for synoptic surface reports, collected four times daily, etc.

These designators would be retained in the proposed systems.

The transmitting stations would be identified by 3-letter CAA location designators, such as DCA = Washington, D.C., LGA = LaGuardia, SFO = San Francisco, etc.

The collection and location designators would be combined to form data program codes which would be recognized by data program selectors located at each station. The data program selector would be common to both the transmitting and receiving sections of each station. The specific devices which would recognize the codes are a series of plug-in devices, each coded to recognize and produce an output for a specific code. These devices are called cards and consist of an "and" gate on a printed wiring board.

When it is time for, let us say the SA collection, the master control station would get on the line and send SA space SA space SA space SA, the date, the time, carriage return, line feed, and a distinctive gate closing signal. This is a nonprinting character which is designated with the nonsense character ▷◁ to avoid confusion. Any transmitting station that had an SA card in its data program selector and also had its tape data selector switch set on SA would cock to transmit. The transmitter operator would set the tape data selector switch to correspond with the type of data he inserts in the tape reader.

All receivers having SA cards in their data program selectors would turn on printers on receipt of SA space. Printers would print

SA SA SA 180231Z (date and time).

The carriage would then return, the line would feed once, and the gate would close on the receipt of ▷◁.

The system would then be ready for the transmission and reception of the SA data collection.

The MC (master control station) would now call up transmitting stations in turn. This would be done as follows:

1. The MC would send SADCA.
2. All receivers with SADCA cards would connect their printers to the line.
3. The DCA (Washington, D. C.) transmitter would cut in and send DCA, its message, carriage return, line feed, and the gate closing signal, ▷◁.
4. All cut-in printers would print DCA and the message, then the carriage would return, the line would feed, and the printer would cut out.
5. When the transmitter sends the gate closing signal it would cut out.
6. When MC sees the gate closing signal it would know that the SA transmission from DCA is completed.
7. MC would send SAXXX (next station

in rotation) and steps 2 through 6 are repeated.

This process would continue until the end of the SA collection. At this point MC would send an end-of-program signal (EP>EP≡≡≡) which all receivers would interpret as a command to shut down the printers.

Table I gives the mode of operation in tabular form.

Detailed Description of the Proposed System

Transmission System

As stated earlier and as shown in Fig. 2 the proposed system would be composed of a backbone route interconnecting hubs with lines branching out to the stations. A typical segment of the proposed system is illustrated in some detail in Fig. 3.

In order to insure more reliable data transmission, the backbone route would consist of two channels, one regular and the other alternate. This is shown in Fig. 4. A level change detector would monitor the channels and automatically switch transmission to the spare in case of trouble, providing the spare were not also in trouble. This switch would only occur on the affected section of the route. Suitable alarms and indications would be provided so that sectionalization for trouble location would be nearly automatic.

The most probable source of noise in a system of this type would be the transmitting loops from the stations. Experience teaches that as few as 100 branch lines or loops can contribute an unacceptable amount of noise. Therefore, since the system might have up to 1,000 transmitting stations, it is clear that noise would present a problem. To limit the effect of noise contributed by transmitting loops, they would all be kept cut off except the one that is actually transmitting data. This would be accomplished by inserting a cut-off circuit in the transmitting leg just before it joins the backbone route.

Data would be transmitted serially on the line by frequency modulation at 1,000 bits per second, which equates to at least 1,000 words per minute. The number of words per minute is governed by the code selected; that is, the number of bits in combination required to form a character and, of course the number of characters in a word. For study purposes, a start-stop 4-out-of-8 code was selected. In this code each character is formed by a unique combination of 8 unit-length pulses, 4 of which are "mark" or "1" pulses and 4 of which are "space" or "0" pulses. Also, for synchronization purposes each character is preceded by a unit-length stop "mark" or "1" pulse. From this it will be seen that each character is composed of 10 unit-length pulses or bits. Also, since there are 5 characters and a space, i.e. 6 characters, per word the following arithmetic obtains:

$$\frac{1{,}000\text{ bits}}{\text{sec}} \times \frac{\text{characters}}{10\text{ bits}} \times \frac{\text{word}}{6\text{ characters}} \times \frac{60\text{ sec}}{\text{min}} = \frac{1{,}000\text{ words}}{\text{min}}$$

Combined Transmitting and Receiving Station

The modulated signals on the line would be detected at receiving stations by a demodulator which converts these signals to serial d-c pulses. This may be seen in Fig. 5.

The d-c pulses are fed into a circuit called a serial-to-parallel converter. This circuit accepts the serial pulses on a single lead and stores them until it has received the correct number corresponding to a character. The converter then spills out the stored bits as simultaneous pulses on the parallel leads, leaving it on the right side as shown in Fig. 5. The converter knows when to do this because the sequence of bits or pulses corresponding to a character is preceded and succeeded by start-and-stop pulses of unit length.

The parallel leads which are hot, i.e., carry pulses, in combination with those which are not hot, i.e., carry pulses, in combination with those which are not hot represent the transmitted character. Later it will be seen that this is the form in which the character was first presented to the system for transmission.

Data Program Selector

The parallel leads feed into the data program selector and station control logic. This circuit examines the leads as they become hot in various combinations for proper sequence of occurrence and uses this information to drive the logic.

For example, let it be assumed that Fig. 5 is the station at Washington, D.C., and that it has been programmed to transmit SA data when called up. When the MC transmits SADCA the parallel leads into the selector become hot in successive combinations which correspond to this, and only this, sequence of characters. The selector recognizes SADCA as a command to transmit SA data and gives information to the control logic, which in turn makes hot a lead corresponding to SA. If the tape data selector switch is

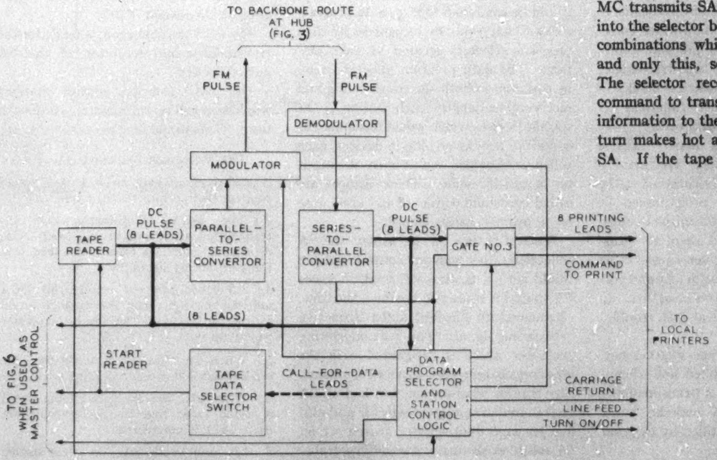


Fig. 5. Combined transmitting and receiving station

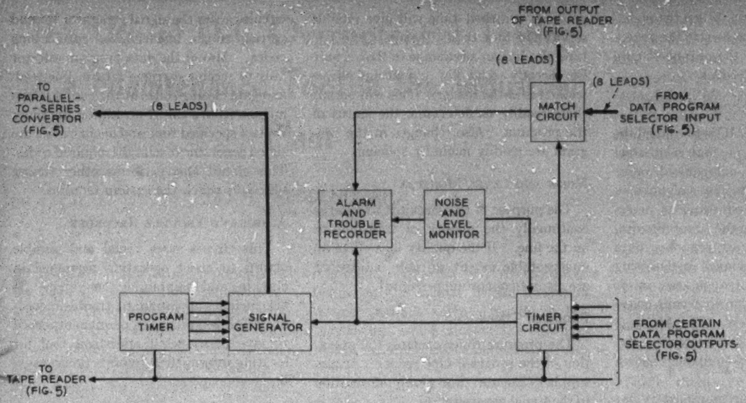

FROM OUTPUT OF TAPE READER (FIG. 5)
(8 LEADS)
(8 LEADS)
FROM DATA PROGRAM SELECTOR INPUT (FIG. 5)
MATCH CIRCUIT
TO PARALLEL-TO-SERIES CONVERTOR (FIG. 5)
(8 LEADS)
ALARM AND TROUBLE RECORDER
NOISE AND LEVEL MONITOR
PROGRAM TIMER
SIGNAL GENERATOR
TIMER CIRCUIT
FROM CERTAIN DATA PROGRAM SELECTOR OUTPUTS (FIG. 5)
TO TAPE READER (FIG. 5)

Fig. 6. Master control station equipment

set on SA, the tape reader will start to send providing there is tape in the reader. When tape is in the reader it makes hot a lead which feeds back to the control logic where, combined with the information derived from the SADCA card, a lead is made hot to the modulator. This causes the modulator, before the tape is read, to send a momentary burst of high-level tone to open the transmission path to the backbone route through the transmitting loop cutoff circuit indicated in Fig. 4.

As another example of the Data Program Selector and Station Control Logic functions, consider its operation at a station other than Washington, D.C., when the MC sends SADCA. The selectors of stations interested in receiving SA data from Washington will have been programmed to do so by the insertion and coding of an address card that will give an output on SADCA. This output causes the logic to make hot leads, which in turn cause Gate No. 3 of Fig. 5 to open. As indicated in Fig. 5, when Gate No. 3 opens it extends leads from the Serial-to-Parallel Converter to the printers.

Programming the Address Cards

From the foregoing it is seen that the function of the Data Program Selector and Station Control Logic is to receive signals from the MC and to translate these signals into station operations in accordance with a program of instructions given it by operating personnel.

The basic circuit element in this program is a plug-in device mentioned earlier known as an address card. All address cards are physically the same. Each card has several inputs but only one output, and all inputs must be made hot before the output becomes hot. It will be shown that, by connecting the inputs of the cards in systematic fashion among a group of information-bearing leads, a unique output can be developed for every hot combination of these leads even though all cards are physically the same. Plug-in cards are used so that selector capacities may be readily fitted to station needs.

Consider a typical station. The operating personnel would know what data they wish to receive and what data they would be required to send. They would accordingly program their particular selector to accomplish this by connecting card inputs to leads which will become hot in response to codes received from the MC, and connecting card outputs to functions in the control logic, which in turn cuts in the printer or turns on the transmitter as required.

For example, assume as before that it is desired to receive the SA data from DCA. An input to the card is connected to a lead in the selector which becomes hot only when S is the first character in an address. Another input to the card is connected to a lead which only becomes hot when A is the second character in the address. This process is continued for DCA. As previously stated, a characteristic of the address card is that it will not produce an output unless all input leads have been made hot. Thus this particular card can produce an output only when the selector receives the address SADCA. In like manner, other cards may be coded to give unique outputs in response to other address and control codes.

The programming of the selector then reduces to coding a card for each data selection it is desired to make. It is probable that the greatest part of a program at a given station will remain fixed. Therefore, the most economical equipment arrangement would be to wire in the initial program. To permit nontechnical operating personnel to make changes in the program as desired, a number of cards, say 100, with patchable inputs and outputs, can be provided. With this arrangement momentary changes in the program could be accomplished quickly. A change in the wired program presents no problem except that it requires some technical skill. It would, of course, be possible to make the whole program patchable, but the cost of such an arrangement and the small additional flexibility thereby achieved makes this look unattractive.

High-Speed Page Printer

The Burroughs matrix page printer is recommended for use in this system because it is inexpensive, can operate in a range from 75 words up to 2,000 words per minute, and is closer to the production stage than any other suitable printer. Because the printer can operate at line speeds, the need for buffer storage at receiving stations disappears, which is a considerable advantage.

Tape Data Selector Switch

The purpose of the tape data selector switch shown in Fig. 4 is to permit the station operator to insert, let us say, an SA tape in the reader with the assurance that it will not be transmitted during an unscheduled or some other scheduled transmission. For example, at the Washington, D.C. transmitter this is accomplished by bringing the outputs of the SADCA, SIDCA, etc., cards, which start the reader, through contacts on the switch designated SA, SI, etc. When the operator inserts the SA tape he sets the switch on SA and thus only the output of card SADCA can start the reader.

Master Control Circuit

The function of the MC is to control the operation of the system, that is, to start transmitting stations at required intervals, monitor the data transmissions to discover and correct omissions of necessary control codes in the transmissions, record for maintenance purposes data which will assist personnel in detecting malfunctioning lines or apparatus, interrupt reguar transmission to permit emergency transmissions, etc. The MC is a combined transmitting and receiving station plus the additional equipment shown in Fig. 6.

The MC would be programmed by operating personnel to call automatically for transmission at desired intervals. The actual programming operation is no more than writing a schedule of collections (transmissions) as is done in the present weather system and then preparing a tape to conform with the schedule.

For example, a tape prepared for the SA schedule would be prepared as follows: SA space SA space SA space SA space, carriage return, line feed, gate closing signal, SADCA, SALGA, SA...etc.; ...EP>EP═══ (end-of-program signal). The program tape is then put in a reader which is controlled by a program timer. The tape is advanced and read under control of the timer and logic circuitry and is stopped at appropriate points for transmissions of date and time group signals by the logic.

A program equivalent to the combined schedules of Services A, C, and O can be put in less than 50 feet of punched paper tape. Because the tape will make but one pass a day, it is expected that a high grade of punched tape will give reliable service for more than 10 years. Punched tape also has an advantage in that it may be directly read by operating people while it is in the reader. This will permit them readily to determine the status of the program. Also, changes in the program are readily made by splicing.

Noise and Level Monitor

The purpose of this circuit is to observe continually the quality of transmission on the line. If the quality falls off to an unacceptable extent suitable indications are given to operating personnel.

Program Timer

The program timer controls the operation of the program tape reader. It also supplies date and time group information to the signal generator.

Signal Generator

This circuit generates date, time group, carriage return, line feed, and gate closing signals when required to do so.

Match Circuit

The purpose of this circuit is to check that called up transmitting stations put their data on the line. It does this by reading and storing the called address (DCA for example) from the tape reader and matching this against the address which the called station transmits. This latter address is read from the input to the data program selector. Failures to match cause appropriate trouble records to be made.

Timer Circuit

The timer circuit times certain operations and causes appropriate signals to be sent as required. For example, if a code match is not made in a specified time, this circuit causes the signal generator to send carriage return, line feed, and gate closing codes. Also, if the data program selector fails to receive carriage return, line feed, and gate closing codes at the end of transmission from a called-up station the timer waits a specified time and again causes the signal generator to send the required codes. This circuit also performs other timing functions which the system requires.

Alarm and Trouble Recorder

This circuit gives visual and audible alarms to assist operating personnel in checking and maintaining the system. It also includes an automatic trouble recording which will provide a permanent record to give personnel a statistical tool for locating intermittent malfunctioning system components.

Appendix. Atmospheric Equations of Motion

$$\frac{du}{dt}=-\frac{1}{\rho}\cdot\frac{\partial\rho}{\partial x}+fv \tag{1}$$

$$\frac{dv}{dt}=-\frac{1}{\rho}\cdot\frac{\partial\rho}{\partial y}-fu \tag{2}$$

$$\frac{dw}{dt}=-\frac{1}{\rho}\cdot\frac{\partial\rho}{\partial z}-g \tag{3}$$

$$\frac{d\rho}{dt}=-\rho\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}\right) \tag{4}$$

$$\frac{dp}{at}=\frac{\gamma p}{\rho}\cdot\frac{d\rho}{dt}=-\gamma p\left(\frac{\partial u}{\gamma x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}\right) \tag{5}$$

x = positive distance toward the east
y = positive distance toward the west
u = eastward component of wind velocity
v = northward component of wind velocity
p = pressure and the partial differential of p with respect to x and y, the pressure gradients
ρ = air density
f = coriolis acceleration: $f=2w(\sin\varphi)$
w = angular velocity of earth's rotation
φ = latitude

LATE DISCUSSION

Discussion of "The Design of the Corner-Reflector Antenna," by H. P. Neff, Jr, and J. D. Tillman, published in Communication and Electronics, July 1959, pages 293–95.

James R. Wait (National Bureau of Standards, Boulder, Colo.): In discussing the ideal corner reflector consisting of two conducting planes which intersect along some apex, the authors state that "the only situation which is amenable to mathematical analysis is that in which the conducting planes are infinite in extent, and in which the angle between the planes is π/n, with n being any positive integer." I wish to draw attention to the fact that a general analysis is available which is valid for any corner reflector angle.[1,2] In fact, it is a series of precisely the same form as their equation 4. When the angle is specialized to π/n, it may be converted to a series of images.

References

1. Electromagnetic Radiation from Cylindrical Structures (book), James R. Wait. Pergamon Press, Inc., New York, N. Y., 1959, chap. 3.

2. Theory of an Antenna with an Infinite Corner Reflector, James R. Wait. *Canadian Journal of Physics*, Ottawa, Ont., Canada, vol. 32, May 1954, pp. 365–71.

Vol.
Pt. 1
1959

The Use of D-C Overpotential Testing as a Maintenance Tool in the Industrial Plant

W. A. WEDDENDORF
MEMBER AIEE

ELECTRIC apparatus represents a large portion of the equipment required for productive activities in the industrial plant: generators to produce power; cables, transformers, and switchgear for transmission and distribution; and motors for transforming electric energy into mechanical power. Insulation, the critical part of electric apparatus, has a limited service life, a variable depending upon the deteriorating and contaminating factors associated with the operation of the equipment. Production interruptions and unnecessary damage from insulation failures represent an economic waste in industry. Needless replacement of insulation, primarily on the basis of age and visual inspection, without substantiating data of its being in questionable condition, further adds to economic waste. As an obvious deduction, there is a definite need for a test which will reliably evaluate the condition of insulation for continued service.

Considerations

An ideal maintenance test would express a definite difference between good, doubtful, and unsatisfactory insulation without reducing the over-all service life or having destructive effect on questionable insulation. To further complicate these utopian requirements the test should be simple to conduct and evaluate and should be one which could be performed by maintenance personnel. The test equipment should be small and ruggedly built so as to be portable and relatively inexpensive. At present there is no known single test method which can fulfill these idealistic requirements.

D-c overpotential testing seems to offer a promising start towards this goal in that the test is simple to conduct, and in recent years considerable experience has been gained in its use better to evaluate insulation. In sensing the demand for d-c test equipment, instrument manufacturers have marketed equipment with improved portability and greater ease of operation. While it is not the author's intention to present d-c overpotential testing as the answer to the insulation testing problem, there is no doubt but that its use would improve the effectiveness of most preventive maintenance programs.

D-C Overpotential Testing

The use of d-c for acceptance testing of new cable installations and subsequent periodic maintenance proof testing has been favorably received by industry. The a-c to d-c conversion factors, representing the differences in the searching qualities between an a-c and a d-c potential, have been established for the various cable insulating materials. Also, there is no question as to what percentage of factory test voltage should be applied for acceptance and proof test values. The duration of the test has, likewise, been established. These details, as related to suggested field testing, are practically spelled out in the general specifications of the IPCEA (Insulated Power Cable Engineers Association) and The Association of Edison Illuminating Companies. Within the proposed test values there should be no fear of detrimental effect on cable insulation.

In reference to the use of d-c overpotential methods in testing the winding insulation of rotating electric equipment, these details are not nicely tied together in a common package. The figures generally used to develop the periodic maintenance proof test value are an a-c to d-c conversion factor of 1.6 and an age factor of 0.65 applied to the a-c factory test of twice-rated voltage plus 1,000. Some engineers have found the 1.6 factor to be too low to develop equal searching qualities with d-c for certain types of insulations, suggesting a conversion factor of 1.8.[1] Others feel that their work in the laboratory has developed data to show that application of a d-c potential, such as used in the voltage-step method, may be detrimental to insulation.[2] With full appreciation of the concern from some quarters towards the general use of d-c overpotential testing, the author, nevertheless, is firmly convinced from a practical point of view that it has a definite application in the maintenance programs of industrial plants. In reality, the gain realized from its expanded use in industry demonstrates that d-c overpotential testing is here to stay, at least until replaced with something better. All engineers should investigate what d-c testing has to offer and make use of this new tool in improving the effectiveness of his maintenance operations.

Much has been written in recent years on the subject of insulation testing by d-c methods.[3] The referenced technical publication offers a complete bibliography and general coverage of insulation testing by d-c methods. It should be of interest in knowing there are groups within the Insulation Subcommittee of Rotating Machinery working on: 1. high-voltage d-c for maintenance testing, and 2. high-voltage d-c as an acceptance test. As a result of this work the AIEE may be in a position in the near future to recommend a uniform procedure for conducting d-c testing of rotating electric equipment. In the meantime, those interested in using this form of testing must establish their methods on the basis of experience, manufacturers' recommendations, and the work of others in this field.

Unfortunately, there have been limited data published on d-c testing relating to the type of equipment found in the industrial plant. The majority of the technical papers have been in connection with the larger high-voltage apparatus in the utility class. Obviously, the smaller equipment is equally as important to the operation of the industrial plant as the larger units are to the utility. Possibly, this brief résumé of the author's experience in this field will stimulate the interest of others in developing a better understanding of its application as well as to its limitations.

Testing of Primary Distribution Cable Circuits

Cable manufacturers high-voltage test all cable before it leaves the factory for assurance of quality construction. It would then be reasonable to assume that cable failures within a few years after installation are primarily due to: 1. damage

---

Paper 59-311, recommended by the AIEE Electrical Insulation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Southeast and South Central District Meeting, Atlanta, Ga., April 8–10, 1959. Manuscript submitted February 5, 1959; made available for printing April 16, 1959.

W. A. WEDDENDORF is with Mutual Boiler and Machinery Insurance Company, Waltham, Mass.

Table I. High-Voltage D-C Field Tests for Cable Installations

(Ozone-Resistant Rubber-Insulated Cable)

Volts Cable Rating	Conductor Size American Wire Gage or MCM	Wall Thickness (Inches)	Basic Field Test Voltages: Upon Installation	Basic Field Test Voltages: Subsequent Proof Tests
600	1–0000	5/64	18.0 kv	13.5 kv
	225–500	6/64	20.5 kv	15.5 kv
	525–1,000	7/64	24.0 kv	18.0 kv
	over 1,000	8/64	27.5 kv	20.5 kv
5,000	8–0000	10/64	33.5 kv	25.0 kv
	225–1,000	11/64	37.0 kv	28.0 kv
	over 1,000	12/64	40.5 kv	30.5 kv
15,000	6–1,000	19/64	63.5 kv	48.0 kv

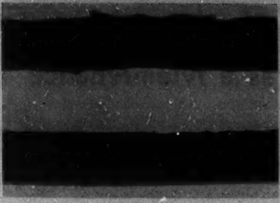


Fig. 1. Insulation failure of 5-kv rubber-insulated cable from sheath currents

in shipment or installation, 2. faulty installation, or 3. improper application. The plant engineer's first line of defense in assuring reliability of operation would be that of conducting a d-c acceptance test immediately after the cable is installed at 80% of the a-c factory test, using the proper a-c to d-c conversion factor for the particular type of insulation. At periodic intervals during the life of any vital cable circuit it could be maintenance proof tested at 60% of factory test for assurance of continued reliability of operation.

Table I gives suggested acceptance and proof test values of the common size and voltage ratings of ozone-resistant rubber-insulated cable used in industry. It should be of interest to know that a 600-volt 525–1,000-MCM (thousand circular mils) rubber-insulated ozone-resistant cable is factory tested at 10 kv a-c. This means that an acceptance test of 24 kv d-c for a period of 10 minutes could be applied immediately after installation. This proof test value is derived by using the 80% field test factor and an a-c to d-c factor of 3.0 for ozone-resistant rubber. Too frequently an important 600-volt cable circuit is placed in service after applying a 500-volt insulation resistance test. Obviously, the 500-volt test will not detect abnormalities which could cause service failures at a later date.

Many distribution circuits in industrial plants are of the 5-kv insulation class. In the conductor size 225–1,000 MCM, the recommended acceptance test would be 37 kv d-c. Along the same lines, the suggested periodic proof test would be 28 kv d-c. Again, the conventional 500- or 1,000-volt insulation resistance test would be far from adequate in assuring reliability of operation. Most cable manufacturers encourage acceptance testing before energizing a new cable installation in assuring reliable installation of a quality product.

Table II. Periodic Proof Testing of Cables

Related Circuit Voltage	Recommended D-C Test Voltage
0–1,000 volts	
#8 up to 500 MCM	10 kv
Over 500 MCM	15 kv
3,000 volts	20 kv
5,000 volts	25 kv
8,000 volts	35 kv
15,000 volts	50 kv

Frequently, the exact construction of the cable in an existing installation is not known, in which case the recommended d-c proof test value is based on the rated circuit voltage. Table II gives the suggested proof test on basis of circuit voltage.

Cable Failure From Improper Application

The 5-kv rubber-insulated cable of Fig. 1 failed from sheath currents due to an improper application for this nonshielded cable. The plant distribution system was comprised of several underground circuits operating at 4.16 kv. The same type of cable was used throughout and had been in service 5 years. A d-c proof test at 28 kv was suggested as a means of determining if other circuits were developing weakened areas from sheath currents. The employees were reluctant to use the suggested test value, so they elected to try a 10-kv test. The leakage current-time curve on the first circuit tested at 10 kv reflected a drooping characteristic shape, typical of good insulation, as shown in lower curve of Fig. 2. On retesting at 15 kv, the leakage current increased with time of voltage application, causing the test to be stopped after 4$^3/_4$ minutes for fear of a breakdown in insulation. This characteristic is shown in upper curve of Fig. 2.

Testing of other circuits reflected unsatisfactory conditions leading to a decision to replace all cable with that more suitable for the application. After pulling the single conductor cable from the ducts, areas of damage from sheath currents were found to verify the interpretation of the test data. Obviously, insulation resistance testing below 10 kv would have given false assurance of reliability of operation.

Proof Testing of Cable Circuits

In another plant the 4.16-kv distribution system was comprised of 13 underground circuits of 3-conductor 5-kv vcl (varnished-cambric lead-covered) cable in sizes ranging from 250 to 600 MCM. A metal-clad switchgear installation was made in 1951, about 500 feet from the old distribution center. The original circuits which had been in service for many years, were spliced and extended with new cable. A recent cable failure gave reason to question the reliability of the distribution system. This apprehension led to a recommendation to d-c proof test all primary circuits, conductor to ground at 16 kv.

The cable had conductor and belt insulation thickness of 0.109 inch and

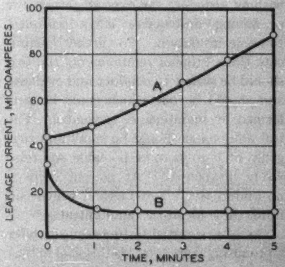


Fig. 2. Leakage current-time curves from high-potential d-c testing of 1,500-foot 4,160-volt cable circuit. 5-kv single-conductor 500-MCM nonshielded ozone-resistant rubber insulation

A—15-kv test
B—10-kv test

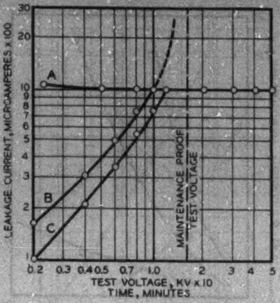


Fig. 3. Leakage current-voltage and leakage current-time curves from d-c overpotential proof test of 1,200-foot 3-conductor 5-kv vcl cable, 250- and 600-MCM, run underground in duct

A—Leakage current-time characteristics
B—Phase C leakage current-voltage characteristics
C—Phases A and B

0.063 inch, respectively, dictating a factory test between conductors at 23 kv a-c and between conductors and sheath at 58% of this value (13.3 kv), in accordance with the IPCEA specifications. The 16-kv d-c proof test conductor to ground was developed by applying the 60% service factor and a 2.0 conversion factor (for varnish-cambric insulation) to the 13.3-kv factory a-c test. Of the 13 circuits tested, five were shown as being unreliable, two in questionable condition, and six as satisfactory for continued operation.

Fig. 3 shows a plot of the leakage current characteristics for one of the circuits termed as being unreliable. The test was stopped at 10 kv on *C* phase and discontinued at 12 kv on phases *A* and *B*, because of the high leakage current of 1,000 microamperes. The steep slope of the curves was an indication that the test voltage could be approaching the breakdown value of the insulation. There was, however, no immediate apprehension of an unintentional breakdown of insulation as far as the test was carried, since the leakage current did reflect a slight drooping characteristic with time of voltage application. Through the use of d-c overpotential proof testing the insulating structure of the primary distribution circuits was evaluated in a nondestructive manner. All 13 circuits were returned to service with a suggestion that the cause for the unsatisfactory test results be investigated and corrected at the first convenient opportunity. The cause could be developing within the insulating structure. Operating conditions did not permit breaking down the weakened areas, thereby developing the cause of the high leakage current at the time.

Many examples could be cited from the author's experiences to show where d-c acceptance and proof testing of cable circuits was instrumental in preventing production interruptions associated with cable failures. Acceptance testing at the suggested values will most definitely ferret out weak insulating structures at splices and terminations which, if not detected, could lead to trouble. A particular mill was experiencing one or more cable failures a year on their 2,400-volt underground distribution feeders. The system was d-c tested at 23 kv (proof test for 3-kv ozone-resistant rubber insulation), breaking down the weakened areas which were found to be in the splices and terminations.

The point of emphasis is that these potential accident sources were corrected at a convenient time without incurring loss of production and at negligible cost. The expense involved was the cost of testing, the electrician's time, and materials required to remake many splices and terminations. The process in this mill was the manufacture of synthetic fibers where continuous power is very critical to plant operations. This testing was accomplished about 3 years ago, and it is encouraging to note that there has not been one case of cable trouble since that time. Others have found high-voltage d-c testing to be of considerable value for acceptance testing of new cables and routine proof testing of existing installations.[4]

Breakdown Test on Generator Winding

In the initial stage of extending d-c overpotential testing to insulation of rotating electric equipment, the writer conducted a breakdown test on a vertical 1,500-kw 2,400-volt 240-rpm water turboalternator. The 18-year old-class *A* insulated stator winding was to be replaced as preventive maintenance. One coil which recently failed at the edge of the slot from mechanical abrasion due to a loose lamination had been cut out of the electrical circuit.

The first step of the test was that of opening the Y connection and proof testing each phase independently at 6 kv with the other two phases grounded. The 6-kv maintenance proof test was developed by applying a 65% age factor and a 1.6 a-c to d-c conversion factor to the factory test value of twice rated voltage plus 1,000. Each phase developed very similar leakage current-voltage values which were linear throughout the 6-kv test range. A plot of these values is shown in the lower curve of Fig. 4.

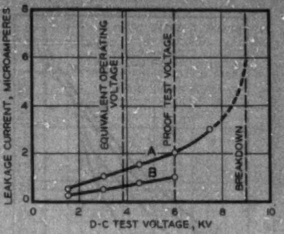


Fig. 4. D-c overpotential breakdown test on stator winding of 1,500-kw 2,400-volt 240-rpm vertical water turbogenerator

A—Breakdown test of complete winding
B—Proof test of each phase independently

The connection was made and the complete winding was tested to breakdown. The insulation of phase 2 broke down after reaching the 9-kv voltage step, in the area adjacent to where the one coil had previously been cut out. These test values are plotted in the upper curve of Fig. 4. The curve was linear through the 6-kv test but reflected a pronounced rise at 7.5 kv. It is evident from examining the curve that the test could have been stopped at 7.5 kv and the 9.0-kv breakdown predicted by extending the curve. This would probably have been more pronounced if smaller incremental voltage steps were used.

Phase 2 was isolated and the remaining two phases tested to breakdown in 1,500-volt incremental steps. Breakdown occurred in phases 1 and 3 on increasing the voltage to 15 kv. The steady state of leakage current were plotted as they were in Fig. 4, and again it appeared that the test could have been stopped at 13.5 kv and the 15-kv breakdown predicted by extending the curve. Probably, if the coils in the weak area of phase 2 could have been replaced without damage to other insulation, the winding would have had several more years of reliable useful life.

It would be well to note: 1. the linear characteristic of the leakage current in both curves of Fig. 4 through the 6-kv step, and 2. the high breakdown value in the unweakened phases, 1 and 2, of better than two times the suggested proof test value. It would seem, from a prac-

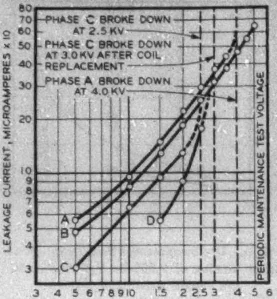


Fig. 5. Periodic d-c overpotential test on 9-year old stator winding insulation of 3,500-kva 2,300-volt 150-rpm vertical water turbogenerator

A—Phase A
B—Phase B
C—Phase C
D—Retesting of phase C

tical point of view, that the 6-kv periodic maintenance test would not be detrimental to this class *A* insulation structure designed for 2,400-volt operation.

Proof Test on Generator Winding

The stator winding of a 3,500-kva 2,300-volt 150-rpm water turboalternator experienced six coil failures from various causes during a 9-year period following rewinding. D-c proof testing was suggested as a means of evaluating the condition of the insulation for continued operation.

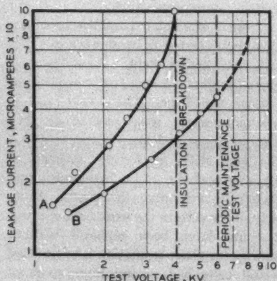


Fig. 6. Periodic d-c overpotential testing of overaged stator winding insulation on large 2,400-volt synchronous motors

A—1,050-hp 2,200-volt 600-rpm synchronous motor
B—2,100-hp 2,300-volt 514-rpm synchronous motor

The Y connection was opened and each phase was d-c tested independently by the voltage-step method in increments of 500 volts. The leakage current-voltage test values are plotted in Fig. 5. Phase *A* broke down without warning at 4 kv. The leakage current at 3.5 kv was 440 microamperes, showing an insulation resistance of about 8 megohms.

Phase *B* withstood 5 kv, the value agreed upon prior to conducting the test as being adequate to detect any defect which might cause an early insulation failure. The hydroelectric station, consisting of five similar machines, operates with a grounded neutral, limiting voltage above ground to about 1.3 kv. The 5-kv proof test value was considered to have adequate searching qualities for this application.

Phase *C* broke down without warning at 2.5 kv. The insulation failures in both *A* and *C* phases occurred at the edge of the coil slot. The coils which broke down were replaced. On retesting, phase *C* broke down at 3.0 kv, which could have been predicted if the test had been discontinued at 2.5 kv and the leakage current retest curve of Fig. 5 extended. The bad coil was replaced, and before the machine was returned to service, all three phase windings had withstood the 5-kv d-c breakdown test. The stator windings of the four other similar machines satisfactorily withstood the 5-kv d-c voltage-step test without incident.

The generator has now been in operation for more than a year since the weak coils detected by d-c testing were replaced. The machine has operated without incident during this period which is somewhat of a record, over previous years of operation. The present plans are to retest all five machines periodically. The testing is to be accomplished during periods of low water so that any weak coils which may break down can be replaced with no loss of power generation.

Evaluation of Motor Winding Insulation

The reliability of two large, slow-speed synchronous motors was questioned, primarily due to the age of the insulation, which was in excess of 25 years. One motor was rated 1,050 hp (horsepower) 2,200 volts, 600 rpm, and the other, 2,100 hp, 2,200 volts, 514 rpm. Both windings were of class *A* insulation. The stator winding construction of these two machines was very similar to that of the water turboalternators in the two previous cases; hence, this might be considered a comparable application for the use of d-c testing. A plot of the test data for these two motors is shown in Fig. 6.

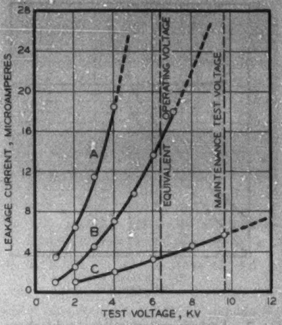


Fig. 7. Plot on cross-sectional paper of periodic d-c overpotential testing of 34-year-old insulation of 7,500-kw 4,000-volt 1,500-rpm turbogenerator

A—Test of April 1957
B—Test of January 1958
C—Test of April 1956

The stator winding insulation of the 2,100-hp motor withstood the 6-kv test. The final steady-state leakage current of 45 microamperes gave an insulation resistance of 130 megohms. The good insulation resistance value, plus the fact that the slope of the leakage current-voltage curve was not excessive, indicate reasonably good insulation. This motor can be continued in operation; however, d-c proof testing on an annual basis to develop trend in the insulation was suggested for greater assurance of safe continuous operation.

The stator winding insulation of the 1,050-hp motor failed at 4 kv. The tests on these two motors were conducted by an engineer from the local service shop. The data were furnished to the author for study and comment. It would appear the test on the 1,050-hp motor could have been stopped at 3.5 kv and the 4-kv breakdown predicted by extending the leakage current curve. The steepness of curve is forecasting warning of an imminent failure of weak insulation. In this manner the stator winding could have been evaluated by d-c testing as being unsafe for continued operation without damage to the insulation. The test was conducted at an opportune time and, although the insulation broke down, the motor was rewound with no loss of production.

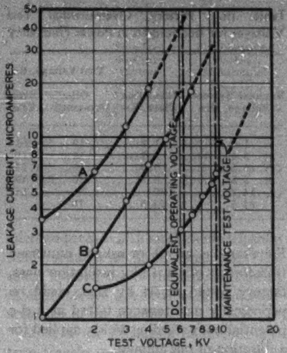


Fig. 8. Plot on log-log paper of periodic d-c overpotential testing on 33-year-old insulation of stator winding of 7,500-kw 4,000-volt 1,500 rpm turbogenerator

A—Test of April 1957
B—Test of January 1958
C—Test of April 1956

Method of Plotting Is Important to Interpretation

A 25-cycle 7,500-kw 4,000-volt 1,500-rpm turboalternator was to be retired from service in 3 years. Plant personnel were apprehensive of reliability of operation in extending the service of 33-year-old insulation. A d-c overpotential test was suggested as a means of evaluating its condition. The machine was taken out of service the night before so as to allow the winding to cool to room temperature. The generator was Y connected with two windings in parallel per phase. The six windings were isolated and each tested independently, recording and plotting the steady-state leakage current for each voltage step. The respective readings were found uniform in value indicating the insulation as being in similar condition throughout the machine. A set of current readings for one winding is plotted in lower curve of Fig. 7, identified as test of April 1956. When plotted on cross-sectional paper the curve was almost linear with moderate slope, an indication of reasonably good insulation. On completion of testing the winding was reconnected and the machine returned to service. An outage of 24 hours without inconvenience to production, plus the small cost of conducting the test, developed assurance of reliable insulation.

The machine had been out of service for several months due to mechanical trouble immediately prior to the subsequent test one year later in April 1957. It was anticipated that the leakage current readings might be higher because of moisture absorption. The values were found in excess of those anticipated, developing an extremely steep curve as shown in Fig. 7, identified as test of April 1957. The test was only carried to 4 kv on each winding for fear of breaking down the insulation. In preparing the report, the data for the April 1956 and April 1957 tests were plotted on log-log paper, as shown in Fig. 8. The shape of the April 1957 curve on log-log co-ordinates materially changes the interpretation. It is evident that the test could have been carried well beyond 4 kv without fear of breakdown. The two vertically separated parallel curves reflect a definite indication of similar quality of insulation, the displacement caused by increased leakage current due to moisture absorption.

An opportunity developed to retest the machine 9 months later, in January 1958. While the machine had not operated appreciably since April 1957, the general atmospheric conditions were less humid; hence, it was expected that the windings would contain less moisture. This observation was correct, as is apparent from the position of the curve in Figs. 7 and 8, identified as test of January 1958.

The author prefers to plot leakage current-voltage curves on log-log paper which tends to present the same relative dimensions of co-ordinates, which aids in the interpretation of slope. Experience would indicate that a rapid increase in leakage current with an increase in voltage in the first part of a test of dry insulation is an indication of deterioration. This is shown in test of April 1956 in Fig. 8 but not apparent in Fig. 7. It is obvious that the method of plotting the test data plays an important function in evaluating the results.

Should the Generator Be Rewound?

The operation of large rotating electrical equipment is normally vital to the production requirements in the industrial plant. Casualty insurance companies have learned from experience that overage insulation in this class of equipment presents an abnormal risk because of the high probability of failure, incurring costly damage to the object in addition to a large production interruption. While everyone recognizes that insulation has a limited service life, age alone should not govern when it should be replaced. In recent years, data from d-c overpotential testing have helped to resolve a difference of opinion between assured and insurer. This is one method of developing tangible data to show the insulation as being reliable for continued service or unsatisfactory to justify the investment of rewinding.

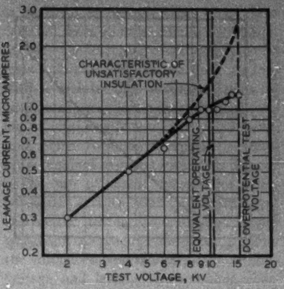


Fig. 9. D-c overpotential testing of 30-year-old stator winding insulation of 5,000-kw 6,600-volt 3,600-rpm turbogenerator

A question developed as to the need of replacing the 33-year-old stator winding of a 5,000-kw 6,600-volt 3,600-rpm turboalternator. The insurance company agreed to defer a recommendation for rewinding if data from a d-c overpotential test indicated the insulation to be in reliable operating condition. It was agreed that the manufacturer's recommended suitability for service d-c test voltage of $2.0E$ plus 800 would be used, developing a test value of 14 kv.[6]

The neutral connection was opened and all three phases were tested independently of each other. The respective voltage-step leakage currents for each phase were uniform in value and very small in magnitude. The recorded leakage current at 14 kv was only 1.2 microamperes, indicating an insulation resistance of 11,500 megohms. A plot of the test values for one phase is shown in Fig. 9. An interesting aspect is that the curve developed a drooping characteristic in approaching the maximum test voltage. Evidently, the small current being measured was not true leakage current but, instead, contained an appreciable absorption component. The explanation is that the absorption component must have been decreasing in magnitude faster than the leakage current was increasing to develop the drooping effect. The situation was aided by a voltage regulation problem inherent to the test set when dealing with small current values. Obviously, the recommendation for rewinding was deferred pending retesting at a later date, preferably on an annual basis.

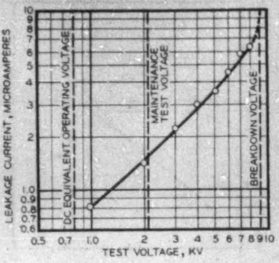


Fig. 10. D-c overpotential breakdown test of 28-year old stator winding insulation of 5,400-kw 480-volt 3,600-rpm turbogenerator

Breakdown Test on 480-Volt Alternator Winding

Permission was obtained to conduct a d-c breakdown test on 28-year old stator winding insulation of a 5,400-kw 480-volt 3,600-rpm turboalternator, which was to be replaced in changing the voltage rating. This was predicted on completing the test in a few hours so as not to interfere with rewinding operations, which placed limitations on what could be accomplished. The connections were opened to facilitate testing each winding independently; however, time was not taken to remove the current transformers from the circuit.

The test voltage was increased in incremental steps of 333 volts until breakdown, which occurred in *A*, *B*, and *C* phases at 13.6 kv, 9 kv, and 11 kv, respectively. Unfortunately, all three insulation breakdowns occurred within the current transformers. While the breakdown strength was not determined, each phase winding comprised of 28-year-old insulation satisfactorily withstood a d-c breakdown test of 9 kv, more than four times the suggested d-c periodic proof test of 2,100 volts. From a practical point of view, these data strengthen the author's conviction that there should be no fear of damage to insulation in reasonably good condition, within the suggested proof test range. The leakage current data developed from the test of *B* phase is shown plotted in the curve of Fig. 10. The curves for phases *A* and *C* were not shown since they would have coincided with *B* phase up to the 8-kv voltage step. A sharp upturn was not present in the leakage current curves to forewarn of an insulation breakdown within the current transformers.

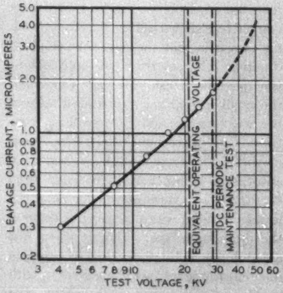


Fig. 11. Periodic overpotential testing of 7-year-old stator winding insulation of 15,000-kw 13.8-kv hydrogen-cooled turbogenerator

Use of D-C Testing in Following Trend of Insulation

The industrial plants have expanded their power generating facilities in keeping pace with increased power requirements to the point where it is not uncommon to find larger hydrogen-cooled generators being utilized. Many of the engineers in these modern plants are using d-c overpotential testing in connection with periodic dismantled inspections to follow the trend of the insulation with years of service. The data plotted in Fig. 11 were developed from testing a 15,000-kw 13.8-kv hydrogen-cooled generator after 7 years of operation. The test of 28.5 kv in accordance with the manufacturer's recommendation, was approached in incremental voltage steps of 4 kv. The leakage current of 1.7 microamperes at 28.5 kv shows an insulation resistance of 16,000 megohms on each winding, phase to phase and phase to ground. This high insulation resistance plus the general shape of the curve indicates very good insulation.

D-C Overpotential Testing Is Applicable to Transformers

Reference 6 suggests maximum values for periodic dielectric tests in the field using both a-c and d-c potentials. The a-c acceptance and periodic proof test values are respectively 75 and 65% of the factory test. For sake of discussion, an abbreviated listing of factory test voltages is shown in Table III.

Both American and National Electrical Manufacturers Association Standards state that "transformers may be subjected to periodic insulation tests using direct voltage from kenotron sets. In such cases, the direct voltage should not exceed the factory test rms alternating voltage."

Table III. Factory Overpotential Test Voltages for Transformers at Power Frequency

Nominal System Voltage, Kv	Insulation Class, Kv	Test Voltage, Kv Oil-Immersed	Test Voltage, Kv Dry-Type
1.2	1.2	10	4
2.4	2.5	15	10
4.8	5.0	19	12
8.32	8.7	26	19
14.4	15	34	31
34.5	34.5	70	

For example, an oil- or askarel-immersed transformer of the 15-kv insulation class, factory tested at 34 kv rms, could be periodically maintenance tested at a d-c potential not exceeding 34 kv, applied for a duration of one minute.

If the d-c test is to be applied to the transformer for a duration of 5 minutes, the value should not be more than 90% of the factory rms test value. In the case of the 34-kv test, the 5 minute duration of the d-c test would be 30.6 kv. The standards recommend that periodic direct-current overvoltage tests should not be applied to transformers of higher than 34.5-kv voltage rating, factory tested at 70 kv.

The plot of Fig. 12 was developed from a d-c test on primary windings of a 3,333-kva 3-phase 13,800/600-volt power distribution transformer. The final test voltage of 30.6 kv was held for a 5-minute duration to check the absorption characteristic.

The transformer had been in service 5 years. The test was requested by a prospective purchaser of the second hand transformer. The leakage current of 3.5

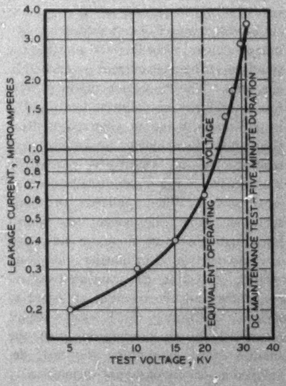


Fig. 12. D-c overpotential test on primary winding of 3,333-kva 3-phase 13.800 Y/600-volt askarel-immersed transformer

Table IV. Test Voltages of Switchgear

Rated Voltage of Switchgear	Standard 60-Cycle Factory Dielectric Test (Kv) (ASA—C37.20)	Suggested Periodic Maintenance Test (Kv)	
		A-C	D-C
2,400	15	11	16
4,160	19	14	20
7,200	26	19.5	27.5
13,800	36	27	38

microamperes at 30.5 kv showed an insulation resistance of 8,700 megohms. The sharp upward rise in the leakage current at about 20 kv was attributed to a characteristic of the askarel.

Most of the insulation failures of dry-type transformers are due to high-resistance carbonized paths caused by leakage currents from foreign deposits. Many plant engineers evidently do not realize that a dry-type transformer cannot be installed and forgotten. They must be cleaned at periodic intervals if a satisfactory life expectancy is to be obtained. The use of d-c overpotential testing, in conjunction with a cleaning and inspection program, will ferret out unsatisfactory insulation. In accordance with the standards, a 15-kv insulation-class dry-type transformer could be periodically d-c tested at 31 kv for one minute, or 27.9 kv for 5 minutes.

Maintenance Testing of Switchgear

Insulation resistance testing of the breakers and bus is generally made during periodic maintenance inspections with a direct reading instrument at 500, 1,000, or 2,500 volts, depending on the practice followed at the particular plant. The readings are generally recorded for comparison, since the trend is more important than a single reading. Insulation testing below operating voltage will primarily reveal the presence of conductive deposits and moisture as opposed deterioration or the developing of a detrimental condition leading to an impending failure. Testing at or above operating voltage has a better chance of detecting a harmful insulation condition which could lead to an insulation failure.

Dielectric tests of switchgear assemblies are made at the factory in accordance with voltages suggested in reference 7. The standard further states dielectric tests may be made at 75% of the standard insulation test values to determine the practicability of continuing equipment in service. This is a low-frequency voltage test for a duration of one minute. Usually it is more convenient to conduct the test with a d-c potential, in which case the general practice is to use a test voltage corresponding to the peak value of the a-c test voltage.

The standards for certain apparatus such as potential transformers may call for a lower test voltage than that suggested. If so, they may be disconnected during test.

Summary

Experience has shown overpotential testing of insulation to be one of the best known methods of verifying that it is in good condition for continued operation.[8] Some engineers suggest this test be made with an a-c potential, since the insulation is stressed with an alternating potential in service, the reason being that there does not seem to be a definite correlation between the searching qualities of the two tests. The author has no fault to find with a maintenance program which incorporates periodic a-c overpotential testing of vital equipment, the point being that some form of overpotential testing is required to detect insulation defects.

The use of direct current for overpotential testing offers definite nondestructive testing advantages. In approaching the predetermined test level by the voltage-step method, leakage current-voltage and leakage current-time characteristics are developed. These data can be interpreted, except in certain circumstances, as indicative of an impending insulation failure or give assurance to proceed with the testing. On completion of the test, much data will have been developed from which an experienced engineer can evaluate the immediate condition of the insulation and follow the trend from subsequent periodic testing.

In offering Accident Prevention Engineering Service to industry, the Mutual Boiler and Machinery Insurance Company has been actively engaged during the past 5 years in using d-c overpotential testing to evaluate the condition of insulation in vital electric equipment. The test projects discussed in this paper were taken from their files and are representative of a general cross section of the work. The early stage of the program was promoted at a time when in the author's opinion the possibility and the effect of an unintentional insulation failure under test was overemphasized as a detrimental factor. In the few instances of unintentional

Fig. 13. Stator winding damage form an insulation failure within a 1,000-kw vertical water turbogenerator

Fig. 14. Stator winding damage from an insulation failure within a 10,000-kw 13.8-kv turbogenerator

breakdowns, the failure usually occurred in the early stage of voltage build up, indicating that the insulation was on the rugged edge of failing in service. Most plant engineers have reconciled their thinking to conclude that it is a definite advantage to have an insulation breakdown at an opportune time under test than to experience a failure in service.

Fig. 13 exhibits severe damage to a 1,000-kw water turbogenerator from an insulation failure in service. The resultant electrical arcing not only made it necessary to replace the stator winding but, in addition, the core structure had to be dismantled and restacked to replace the burned laminations. As is frequently the case, the accident occurred during a period of ample water conditions when the power requirement was critical to production. The damage from electrical arcing, associated with an insulation failure in service, greatly extended the outage time.

The stator winding of the 10,000-kw 13.8-kv turbogenerator of Fig. 14 failed in service from an insulation breakdown, after many years of reliable operation. The power shortage from the outage of this generating unit curtailed plant operations. Repairs were expedited on around-the-clock 7-days-per-week operation in holding the production loss to a minimum. Periodic d-c overpotential testing will materially aid in preventing the type of costly accidents shown in Figs. 1, 13, and 14. Frequently d-c testing will reveal the need of maintenance which, if not corrected, may lead to an insulation failure.

Year by year, plant engineers are becoming more maintenance minded. Realizing that accidents are too costly to afford, they are in constant search for new means of accident prevention. The interest in all phases of plant maintenance is being focused on methods of nondestructive testing. D-c overpotential testing seems to offer the industrial plant's maintenance program a method for approaching a nondestructive evaluation of insulation. In the author's practical experience there has been no evidence to reflect damage to otherwise good insulation as the result of having been subjected to d-c proof testing, nor has there been any case where insulation has failed in service after testing to indicate the d-c voltage did not search out impending defects.

References

1. Techniques and Examples of High-Voltage D-C Testing of Rotating-Machine Windings, C. L. Sidway, B. R. Loxley. *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 72, Dec. 1953, pp. 1121–29.

2. Predicting Insulation Failures with Direct Voltage, C. M. Foust, B. V. Bhimani. *Ibid.*, vol. 76, Dec. 1957, pp. 1120–39.

3. Insulation Testing by D-C Methods. *Technical Publication 22T1*, James G. Biddle Company, Philadelphia, Pa.

4. High Voltage D-C Testing of Cables and Cable Fault Location, B. J. Nankervis. *AIEE CP 56-491* (available on request).

5. Insulation Testing for Armature Windings of A-C Turbine-Generators. *Publication GEI-64108*, General Electric Company, Schenectady, N. Y.

6. Requirements, Terminology, and Test Code for Distribution, Power, and Regulating Transformers and Reactors Other Than Current-Limiting Reactors. *ASA C57.12-1956*, American Standards Association, New York, N. Y., sec. 12-92.040, 1956.

7. Switchgear Assemblies and Metal-Enclosed Bus(AIEE 27). *ASA C37.20-1955*, American Standards Association, Jan. 1955.

8. Maintenance Overpotential Tests for Armature Windings in Service. *Publication GER-310C*, General Electric Company.

Vol.
Pt. 1
1959

Design of Polyethylene-Insulated Multipair Telephone Cable

A. S. WINDELER
NONMEMBER AIEE

CABLES are designed to have special characteristics which will make them suitable for particular uses or purposes. In the case of telephone cable the purpose is primarily to transmit information at voice and carrier frequency during an expected life of approximately 30 years. To achieve this, certain mechanical and electrical characteristics must be built into the cable. Mechanical characteristics include such characteristics as resistance to environmental factors including heat, cold, wind, lightning, moisture, sunlight, and oxygen, as well as resistance to damage by rodents and insects, and by workmen. Other desirable mechanical characteristics involve factors of convenience of manufacture, installation, and joining.

The electrical characteristics include such factors as adequate dielectric strength and freedom from crosstalk and, of course, the all-important transmission properties. It is the purpose here to consider only the design of cable to obtain the desired transmission properties. In the case of multipair cable this reduces to adjusting the dimensions of the conductors and the insulation, so as to obtain the desired resistance, inductance, capacitance, and conductance of the cable pairs. From these four primary constants, the secondary parameters of nonloaded circuits, namely, the propagation constant and characteristic impedance, can be calculated. Moreover, with the primary constant data and the loading coil resistance, inductance, and spacing, the secondary constants of loaded circuits may be obtained.

Cost is, of course, one of the fundamental considerations of any cable design. This involves an evaluation of such factors as availability and cost of the material selected, reliability of supply and probably price trend, serviceability as compared with existing materials, etc. These considerations lie beyond the scope of the present article, which deals specifically with problems of design to obtain the desired electrical characteristics. The problem is then to determine the dimensions which will give the desired primary constants; this means determining the the way in which the primary constants vary as the dimensions are changed.

It is our purpose here to consider only the design of solid polyethylene-insulated multipair cable and to discuss only the voice frequency primary constants.

At voice frequencies the main interest here is in the capacitance and resistance of the pairs; this is because the secondary parameters of nonloaded circuits are dependent only on these constants and the frequency of transmission. In the case of loaded circuits the inductance of the cable pairs plays only a small part in

---

Paper 59-793, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted April 6, 1959; made available for printing May 28, 1959.

A. S. Windeler is with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

The author is indebted to F. W. Horn for helpful discussions, to L. M. Rackson, E. O. Bauer, and E. Schultz for the electrical testing of cable samples, and to Mrs. B. M. Felderman for numerous computations.

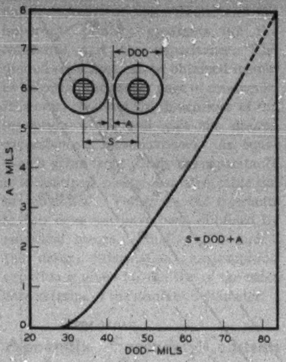


Fig. 1. **Interaxial spacing as a function of diameter over the dielectric**

determining the secondary constants. The effect is quite limited because the inductance of the loading coils far outweighs that of the cable pairs. With polyethylene-insulated cable, conductance has negligible effect in either case. This is true because of the remarkably low power factor of polyethylene. For convenient reference the equations[1] which express the secondary constants in terms of the primary constants are given in the Appendix.

Consider now the problem of determining the voice frequency resistance, inductance, and capacitance from the dimensions of the conductors, and the insulation. For this purpose a representative voice frequency will be taken as 1,000 cycles per second.

At this frequency, for the gages of interest, namely 19, 22, 24, 26 American Wire Gage, the a-c resistance of a cable pair is equal to the d-c loop resistance. For copper wires at 68 degrees Fahrenheit this becomes

$$R=\frac{0.1095}{d^2} \tag{1}$$

where

R = resistance, ohms per mile
d = conductor diameter, inches

Such simplicity is possible only because the skin effect, and certain proximity effects, so important at carrier frequencies, are negligible at voice frequencies for the wire sizes being considered.

For the same reason the equation for the inductance of cable pairs at voice frequencies is very simple, being the same as that of a pair of wires in free space,[2] namely

$$L=1.482\log_{10}\frac{2S}{d}+0.1609 \tag{2}$$

where

L = inductance, millihenrys per mile
S = interaxial spacing of wires
d = diameter of wires

The last term is the so-called internal inductance of the two conductors, which is a constant (0.1609 millihenry per mile) for all sizes.

One question which arises immediately is how to control the interaxial separation S, of the wires of a pair. Fortunately, S is dependent primarily on the diameter over the dielectric, generally referred to as the DOD. This dimension can be contolled quite precisely during the extrusion of the insulation over the conductor. For small values of DOD, normally associated with the finer gages, the value of S is exactly the same as the DOD. This simply means that the insulated conductors fit closely together with no space between them. For larger DOD values, normally associated with the heavier gages, there will be a small gap between conductors, as illustrated in Fig. 1. The extent of this gap depends upon the stiffness of the conductors and upon the tension and compression to which the core is subjected during manufacture and installation. Experience with solid polyethylene-insulated conductors demonstrates that the average value of S for a given DOD can be obtained from the empirical relation expressed by the curve in Fig. 1.

The problem of predetermining the capacitance is more difficult. To compute the pair capacitance one needs not only the dimensions d and S, but also the effective dielectric constant ϵ. The computation is further complicated by the fact that presence of the surrounding wires influences the capacitance of any pair. The surrounding wires form an electrostatic shield, the effective diameter of which is a function of S, the interaxial separation of the wires of a pair. This shield is indicated by the dashed circle in Fig. 2. The effective diameter of this shield is represented by the D. Any pair in the cable thus becomes effectively a shielded pair, the capacitance of which can be calculated by an equation developed a number of years ago by Mead.[3]

$$C=\frac{0.01944\epsilon}{\log_{10}\left(\dfrac{2S}{d}\,\dfrac{D^2-S^2}{D^2+S^2}\right)-0.1086\delta_{12}} \tag{3}$$

where C = capacitance in microfarads per mile, and δ_{12} is a series function of $d/2S$ and S/D. Values of δ_{12} for various values of $d/2S$ and S/D are plotted in Fig. 3.

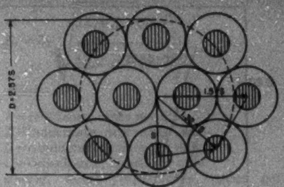


Fig. 2. **Cable pair with surrounding wires as an effective shield**

The relation between S and D may be arrived at by the following reasoning. Consider any pair in the interior of a multipair cable. It may be assumed that the two wires of this pair are surrounded by an oval layer of eight wires as indicated in Fig. 2. It is apparent from the geometry that the distance from the center of the pair to the centers of the two nearest surrounding wires is equal to S, while the distance to centers of the two farthest wires is $1.5S$. The centers of the four remaining wires are $1.32S$ distant from the center of the pair. These eight distances averaged gives $1.285S$ as the mean radius of the layer of wires. As a first approximation this may be considered to be one half the mean diameter D of the effective shield. The S/D ratio then becomes 0.389.

To obtain the dielectric constant ϵ, the remaining parameter involved in equation 3, reference must be made to the measured capacitance of cables for which the ratio of $2S/d$ is known (obtained, as mentioned, from inductance measurements). Data were obtained on eight cables of four different gages in the nominal 0.083-microfarad-per-mile capacitance range; see Fig. 4. The solid curve was calculated with the use of a dielectric constant of 1.80, which seems to fit the data fairly well. The dotted curves above and below the solid curve were calculated with the use of the dielectric constants of 1.90 and 1.70 respectively. These represent the extremes in deviation from the average. It is convenient to lump all variations into the dielectric constant, although some of the deviations may well be the result of variations in the S/D ratio. The S/D ratio for the curves in Fig. 4 was assumed to be constant at a value of 0.389.

The values of the dielectric constant and S/D ratio have been determined independently from capacitance measurements on a number of cables by the method described by Maupin[4] in which

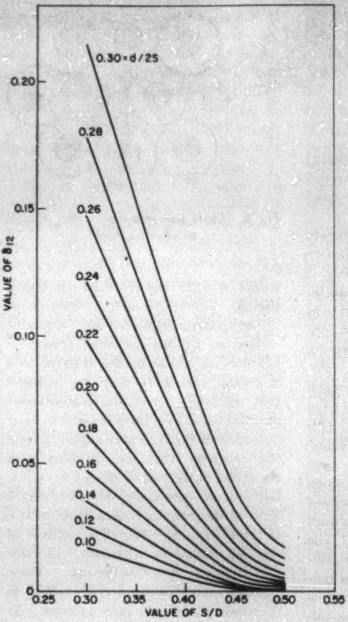


Fig. 3 (left). Value of δ_{12} as a function of $d/2S$ and S/D

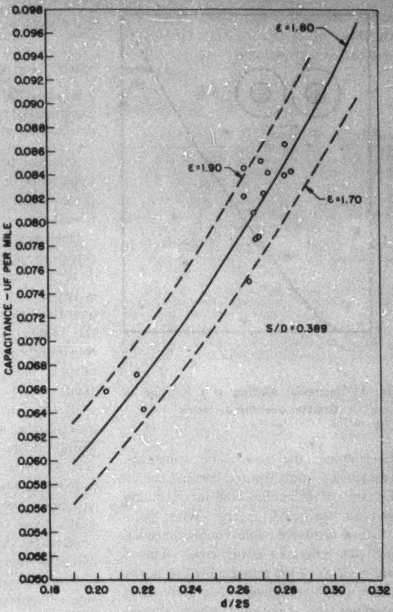


Fig. 4 (right). Capacitance as a function of the ratio of conductor diameter to interaxial spacing

measurements are made of the mutual capacitance and the direct capacitance to ground of cable pairs. These two quantities represented by C_{mut} and C_g respectively are functions of S, d, D, and ϵ, i.e.,

$$C_{mut}=f_1(S,d,D,\epsilon)$$

$$C_g=f_2(S,d,D,\epsilon)$$

Since S can be determined from the inductance measurements and d is known for each gage, the two equations can be solved for the two unknowns D and ϵ. The average value of S/D obtained in this manner is 0.376. Since this method of obtaining the S/D ratio is not very precise, this is considered a good check on the value deduced from geometrical considerations. The average value of ϵ obtained in this manner is 1.83, which is considered a good check on the value of 1.80, obtained from the data in Fig. 4.

It will be noted that ϵ appears to be constant for all values of $d/2S$. For the condition where S equals the DOD (fine gages and small DOD values) one would expect that the ϵ would increase as $d/2S$ is increased, since the ratio of the volume of insulation to the volume of air in the cable increases with an increase in $d/2S$ or an increase in d/DOD. However, the low-capacitance cables (small $d/2S$ or d/DOD ratios) are normally made only in the 19-gage size with the relatively large DOD of 75 mils. Referring to Fig. 1, it can be seen that with such a large DOD value the air space between the insulated conductors of a pair is relatively large (6 mils). This tends to counteract the effect of the small d/DOD ratio. The net result appears to be a constant ϵ over the range of $d/2S$ being considered.

In designing a cable to have a given capacitance one needs only to select the appropriate $d/2S$ ratio from Fig. 4. Knowing the value of d for the gage selected, one can calculate S. From Fig. 1, one can then determine the DOD to which the wire should be insulated.

To complete the design of a cable it is necessary to know one further parameter, i.e., the diameter of cable core to which a given number of pairs will build up. It is common practice in designing cable to allow a theoretical space or area for each pair in the cable. This theoretical area, measured in kilocircular mils (kcm) is known as the space per pair. For instance, it is known that 100 19-gage pairs having conductors with a DOD of 60 mils ($S=63.9$ mils) will form a core 1.17 inches in diameter. The total area involved here is approximately 1.37×10^6 circular mils or 1,370 kcm. This number divided by 100 gives the space per pair, which is 13.7 kcm.

The exact space which pairs of a given size will occupy depends upon the tensions and pressures encountered in the twisting, stranding, and cabling operations. The following empirical equation will, however, serve as a guide in selecting the number of kcm to allow for a pair of a given size, i.e., a given value of S.

$$\text{Space per pair}=3.35S^2\times10^3\text{ kcm} \qquad (4)$$

where S is the interaxial spacing in inches and is related to the DOD as indicated in Fig. 1. Having selected the space per pair one must multiply it by the number of pairs in the cable. This gives the total area of the core in kcm. This area is then coverted to circular inches and the square root extracted to obtain the diameter of the core in inches.

Conclusions

In designing cables to have certain transmission characteristics one must choose the dimensions of the conductors and the insulation so as to produce the primary constants which will give the desired characteristics. In the case of multipair cable at voice frequencies the important primary constants are the resistance and the capacitance. The desired resistance can be obtained merely by selecting the proper gage of conductor. However, choosing the dimensions of the insulation which will give the desired capacitance is more involved. An equation is given from which the capacitance of a multipair plastic-insulated cable can be calculated. Values for the constants in this equation have been obtained by empirical means. Finally, to complete the design information, an empirical equation is given for the size of the cable core in terms of the insulation diameter.

Appendix. Secondary Constants in Terms of Primary Constants

Nonloaded Cable

The propagation constant is

$$P=\sqrt{(R+j\omega L)(G+j\omega C)}$$

$$P\doteq\sqrt{j\omega RC}$$

approximately at voice frequency.

The characteristic impedance is

$$Z_0=\sqrt{\frac{R+j\omega L}{G+j\omega C}}$$

$$Z_0\doteq\sqrt{\frac{R}{j\omega C}}$$

approximately at voice frequency.

Loaded Cable

The propagation constant per mile is

$$P'=\frac{2}{N}\tanh^{-1}\sqrt{\frac{2Z_0\tanh\frac{NP}{2}+Z_C}{2Z_0\coth\frac{NP}{2}+Z_C}}$$

The mid-section characteristic impedance is

$$Z_0'=Z_0\sqrt{\frac{2Z_0+Z_C\coth\frac{NP}{2}}{2Z_0+Z_C\tanh\frac{NP}{2}}}$$

where

R = resistance, ohms per mile
L = inductance, henrys per mile
G = conductance, mhos per mile
C = capacitance, farads per mile
$\omega=2\pi f$, where f = frequency, cycles per second
$j=\sqrt{-1}$
N = number of miles per loading section
Z_C = loading coil impedance

References

1. TRANSMISSION CIRCUITS FOR TELEPHONIC COMMUNICATION, K. S. Johnson. D. Van Nostrand Company, Inc., Princeton, N. J., pp. 154–55.

2. ELECTROMAGNETIC WAVES, S. A. Schelkunoff. D. Van Nostrand Company, Inc., p. 284.

3. SHIELDED CABLE SYSTEM, S. P. Mead. *U. S. Patent no. 2,086,629*, Apr. 14, 1936.

4. MEASUREMENTS IN MULTIPAIR CABLES, J. T. Maupin. *Bell System Technical Journal*, New York, N. Y., vol. 30, July 1951, pp. 652–67.

Magnetic Amplifier Circuits—A Classification of Half-Wave and Full-Wave Nonreversible and Reversible Self-Saturating Circuits

DAVID L. McMURTRIE
MEMBER AIEE

THIS ARTICLE attempts to give an orderly, critical classification of a wide variety of self-saturating single-phase series-control magnetic-amplifier output circuits. It the article is successful, it will be clear that a confusing multitude of circuits appearing in the literature may be fitted into a very simple scheme. The relative merit of various alternative circuits for accomplishing different types of output will also become apparent.

The term "self-saturating" implies the use of a rectifier in series with each reactor winding in the output circuit. The more limited simple-saturable type of circuit is not dealt with. "Single-phase" circuits are circuits in which the supply voltages used are of a single phase. "Series control" results from the fact that the only power sources considered are voltage sources. With voltage sources, the reactor elements must act as variable impedance elements in series with the power source and the load. Where current sources are considered, the reactor elements must be used as variable impedance elements in shunt between source and load. The present treatment deals with output circuits only, omitting the details of control and bias. As is developed in the course of the article, there are important constraints which exist in the output circuitry independent of the manner of control, and it is possible to establish criteria for good output circuits independent of control arrangements. Moreover, once a degree of order has been established for output circuits, it is possible to study control arrangements applicable not just to single output circuits, but to classes of output circuits which will be identical insofar as establishing control and bias are concerned. Control and bias are not dealt with in the present article, however.

The manner in which the various half-wave and full-wave nonreversible and reversible circuits are derived is quite simple. The familiar half-wave (HW) circuit, the simplest form of self-saturating circuit, is used as a basic building block with which to construct the more complex reversible half-wave (RHW), full-wave (FW), and reversible full-wave (RFW) output circuits. The technique, with certain modifications that are explained in course, is a straightforward way of deriving even the most complicated circuits, many of which impress the uninitiated as hopelessly complex affairs arrived at in some mysterious inexplicable manner. The derivation technique also allows a quick search for equivalent or improved circuits which may be more economical in one or more of the components involved; i.e., reactor windings, rectifiers, and supply or load transformers.

From this point on the abbreviations mentioned before will be used, adding "AC" and "DC" where necessary to indicate the particular type of output. Thus an FWDC circuit will be a circuit having full-wave d-c output, and an RHWAC circuit will be one having reversible half-wave a-c output, and so on.

There is one serious stumbling block in the process of building complex circuits

---

Paper 59-790, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted August 21, 1958; made available for printing April 14, 1959.

DAVID L. McMURTRIE is with the Raytheon Manufacturing Company, Wayland, Mass.

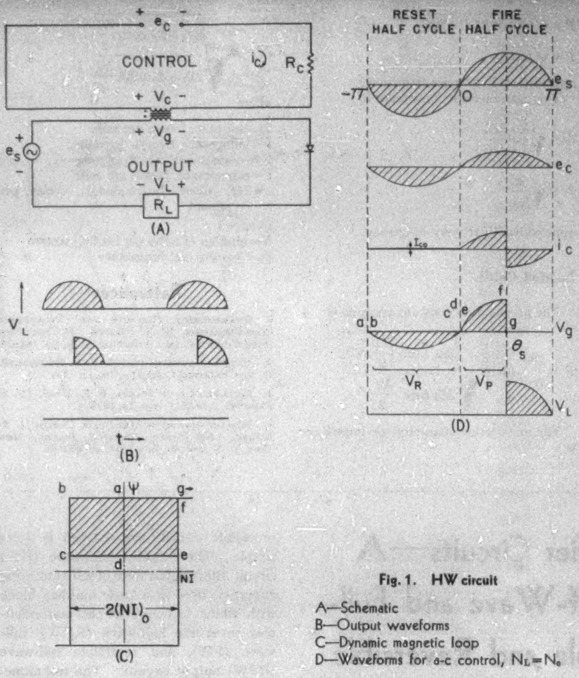


Fig. 1. HW circuit

A—Schematic
B—Output waveforms
C—Dynamic magnetic loop
D—Waveforms for a-c control, $N_L=N_c$

from the HW circuit (or in constructing them in whatever fashion). If there were no such stumbling block, the substance of the present paper would be somewhat trivial, and the ground would have been thoroughly covered long since. The unpleasant fact is that serious interactions often occur when HW circuits are combined to form more complex structures. Anyone who has worked with magnetic amplifiers has run into booby-trapped situations where the seemingly straightforward push-pull (or other) type of circuits were completely inoperable for the purpose intended because of unforeseen interactions. Treated as controllable synchronous switches, the HW circuits should work perfectly well to give the desired over-all circuit performance, but taken as actual reactors with definite flux-resetting requirements, the circuits are unworkable. It is the essence of the present paper that important output circuit interactions are recognized, measured, and classified, thus allowing a meaningful evaluation and classification of the whole range of single-phase self-saturating output circuits.

Nomenclature

e_b = bias voltage
e_c = control voltage
$e_s=E_m \sin \omega_s t$ = supply voltage
i_c = control current
$I_{co}=(NI)_o/N_c$
i_L = load current
N_c = control-winding turns
N_L = load or output winding turns
$(NI)_o$ = coercive ampere-turns
R_c = control-circuit resistance
R_L = load resistance
V_c = control-winding voltage
V_g = gate or load-winding voltage
V_p = prefire voltage $\equiv V_g$, $0<\theta<\theta_s$
$\bar{V}_p$ = prefire voltage, average value over half cycle
V_r = reset voltage $\equiv V_g$, $-\pi<\theta<0$
$\bar{V}_r$ = reset voltage, average value over one-half cycle
V_L = load voltage (reflected to primary in case of autotransformer output)
V_o = output voltage $\equiv 0$ when no output is being delivered, $\equiv e_s$ when output is being delivered (V_o and V_L are synonymous except for circuits using isolating or mixing resistors which consume a portion of the output voltage)
θ_s = core saturation angle in fire half-cycle
ψ = magnetic flux
ω_c = control-signal angular frequency
ω_s = supply or carrier angular frequency

HW Circuit

The simplest form of self-saturating magnetic amplifier is the HW circuit shown in Fig. 1. Knowledge of its manner of operation is assumed. The output and other waveforms are illustated assuming an idealized square-loop magnetic characteristic. An a-c control voltage is shown, but the control could as well be d-c or it could be a combination of the two, including a rectifier to yield the Ramey circuit, etc. The output obtainable is, as shown, nonreversible having

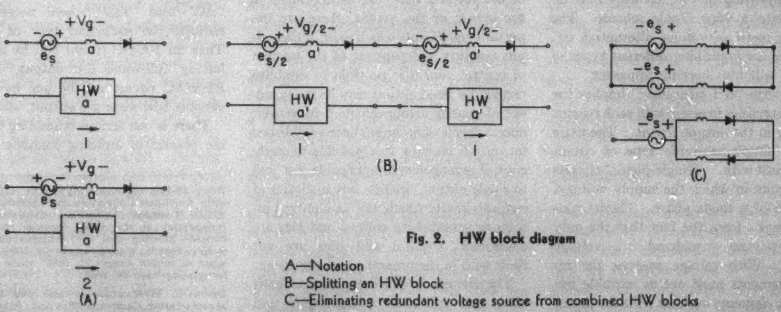


Fig. 2. HW block diagram

A—Notation
B—Splitting an HW block
C—Eliminating redundant voltage source from combined HW blocks

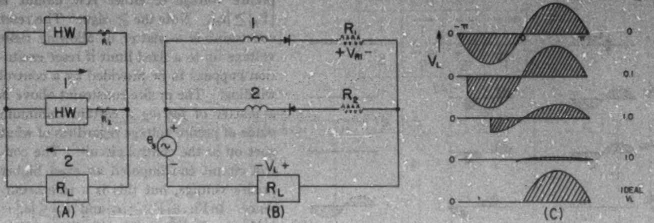


Fig. 3. Back-to-back HW circuits for FWAC or RHW output

A—Block diagram
B—Schematic with redundant voltage source eliminated
C—Maximum HW output versus R_1/R_L

both a d-c and an a-c component. For convenience in later discussion, the gate or load-winding voltage V_g is resolved into two components, timewise, the reset voltage V_r and the prefire voltage V_p.

The block diagram notation which is to be used in constructing more complex circuits is shown in Fig. 2(A). Note that in the scheme used here, the HW block includes a voltage source. The arrow beneath the block indicates the direction of current flow through the rectifier, and the number *1* or *2* is used to indicate the half-cycle during which the voltage source has the proper polarity to produce positive current through the rectifier. Fig. 2(B) shows the notation to be used if the wish is to split an HW block into two equal parts as will be done often in arriving at bridge-circuit equivalents of the more complex circuits. The fact that in such cases the supply voltage and number of gate-winding turns are cut in half will be kept track of by using the prime as shown, and the double prime if the process is repeated. Fig. 2(C) shows the way in which superfluous voltage sources may be eliminated when HW circuits are combined. The process may be repeated several times in the case of multiple-HW-circuit combinations, Figs. 9(B) and (C) being one example.

Interactions Occurring When HW Circuits Are Combined

As mentioned in the opening paragraphs, certain important interactions may occur between individual HW circuits which are combined to yield RHW, FW, and RFW circuits. These interactions may be classified as follows.

Interactions in Output Circuit

Reset voltage constraints. In the basic HW circuit shown in Fig. 1, the magnitude of the reset voltage cannot exceed the supply voltage, i.e., $|V_r| \leq |e_s|$. If one attempts to impose a greater reset voltage, via the control source, current flows in the load circuit causing a corresponding number of ampere-turns in the control circuit. Provided $(N_c/N_L)^2 R_L << R_c$, $|V_r|$ will remain $\doteq |e_s|$, and the excess control voltage will appear across R_c. If R_c is very small, $|V_r|$ can be increased beyond $|e_s|$, but only at the price of large load and control currents with a subsequent loss in gain and efficiency. If two HW circuits are now put back-to-back as shown in Fig. 3, with control voltages applied to control circuitry not shown, it is not difficult to see that the constraint $|V_{r2}| \leq |e_s|$ holds only as long as core *1* is unsaturated and $V_L = 0$. Once either core saturates, it short-circuits the load winding on the opposite core. When core *1* saturates, then, necessarily, $V_{r2} = 0$ assuming ideal rectifiers and zero gate-winding resistances $R_1 = R_2 = 0$. Thus $|V_r| \leq |e_s|$. In terms of average values over a half cycle, $|\bar{V}_{r1}| \leq |\bar{e}_s| - |\bar{V}_L| = |\bar{V}_{p2}|$. Similarly $|\bar{V}_{r2}| \leq \bar{V}|_{p1}|$. In the steady state, $|\bar{V}_{r1}| = |\bar{V}_{p1}|$ and $|\bar{V}_{r2}| = |\bar{V}_{p2}|$. Hence $|V_{r1}| = |V_{r2}|$ which states that the reset voltages on the two cores must be equal. The prefire voltages must also be equal, and the load voltage on the two half-cycles must be equal. This may be called a condition of constrained balance. Considering just the block diagram of Fig. 3(A), it might have been possible to use this circuit for RHW output. For this type of output, one half of the output circuit must be "shut off" while the other half delivers output of one polarity and vice versa for output of the opposite polarity. The constrained balance condition just observed makes it impossible to shut off one half of the circuit while the other is delivering output. The circuit may be operated perfectly well for FWAC output but not for RHW output. As will be observed later, all back-to-back HW and FW circuits working directly into a common load to deliver reversible d-c suffer from this type of constrained balance. Without modification, these circuits are inoperable as reversible d-c output circuits.

One way in which the preceding constrained-balance condition may be modified is by adding the isolation resistors shown dotted in Fig. 3. Now the reset constraint on either core is $|V_r| \leq |e_s| - |V_L| = |e_s| - R_L|V_o|/R_L + R_1$. (As shown in the Nomenclature, a difference is noted between output voltage V_o and load voltage V_L. $V_o \equiv 0$ when no output is being delivered and $\equiv e_s$ when output is being delivered. V_L may be only a portion of V_o when isolating or mixing resistors are used or when output current bypasses the load and "circulates" through the load windings of other saturated cores.) Note that only one of the isolating resistors at a time is effective and that this resistor gains its effectiveness only by passing load current, thereby reducing the load voltage. If an attempt is made to increase the reset voltage (via the control circuit) beyond the limit mentioned, circulating current will flow and the reset may be increased by the voltage drop produced across the load and isolating resistors. However, this increase will be negligible if $R_L, R_1, R_2 << (N_L/N_c)^2 R_c$ and will be costly in gain and power in any event. It is important to notice that

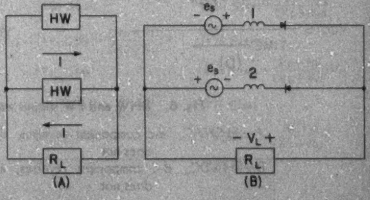


Fig. 4. Back-to-back HW circuit for RHW output
A—Block diagram
B—Schematic

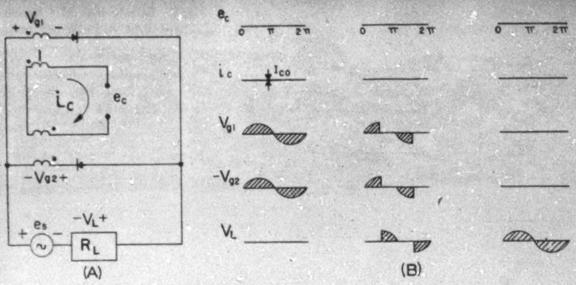


Fig. 5. FWAC circuit with control windings added

A—Schematic
B—Waveforms; any value of output obtained with zero input

even with the use of isolating resistors it will not be possible to shut off one half of the circuit completely while the other half of the circuit is conducting fully (over a complete half-cycle). In order to accomplish this the constraint would have to reduce to $|V_r| \leq |e_s|$, and this is impossible except in the limit $R_1/R_L = \infty$, in which case all the output power appears across R_1 and none across the load. Fig. 3(C) illustrates the situation. In practice, therefore, only a moderately successful compromise can be attained in which isolating resistors rob a sizable portion of the output power and still do not give full output.

Prefire voltage constraint. The reset voltage constraint just dealt with is only half the picture; the other half is the prefire voltage constraint. In the circuit of Fig. 3, as in the basic HW circuit, the prefire voltage of either HW circuit is $|V_p| \geq |e_s|$. Note the $\geq$ sign. The reset constraint is a matter of allowing reset voltage up to a fixed limit if reset excitation happens to be provided by a control winding. The prefire constraint above is a matter of *forcing* a certain minimum value of prefire voltage regardless of what goes on in the control circuit. The control circuit can imposed an even higher prefire voltage, but this is a rare occurrence. In Fig. 3 $|V_p| \geq |e_s|$ and $|V_r| \leq |e_s| - k|V_o|$, $k = R_L/(R_L + R_1)$. As discussed in the preceding section, one of the HW circuits cannot be completely cut off while its mate is fully firing, and full RHW output cannot be obtained, therefore, unless $|V_r| \leq |e_s|$. To this should be added the requirement for keeping one of the HW circuits shut off is that sufficient reset voltage area is allowed across its gate winding so that $|\bar{V}_r| = |\bar{V}_p|$ without forcing $|\bar{V}_p|$ to be reduced by saturation of the core before the end of the half-cycle. It is not always true that the prefire constraint is $|V_p| \geq |e_s|$. The basic criterion is that the maximum allowable value of $|V_r|$ be equal to or greater than the minimum mandatory value of $|V_p|$ when the core

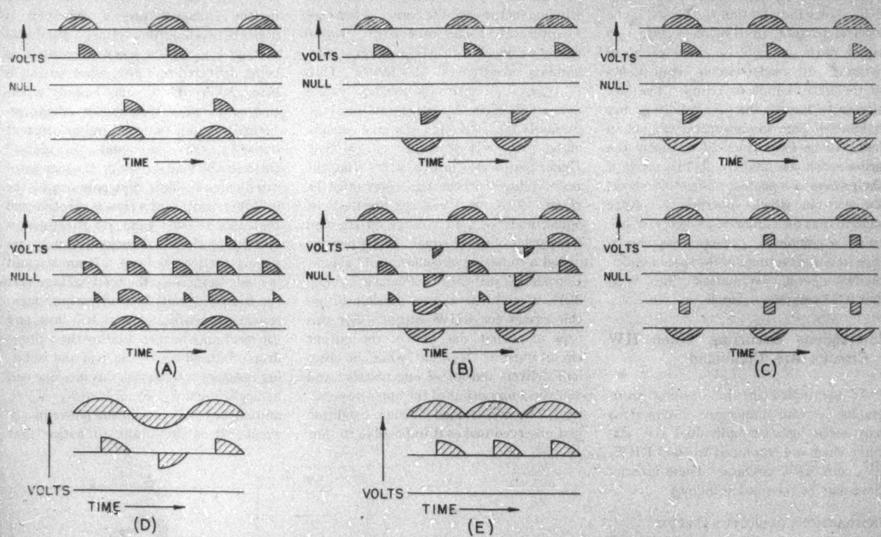
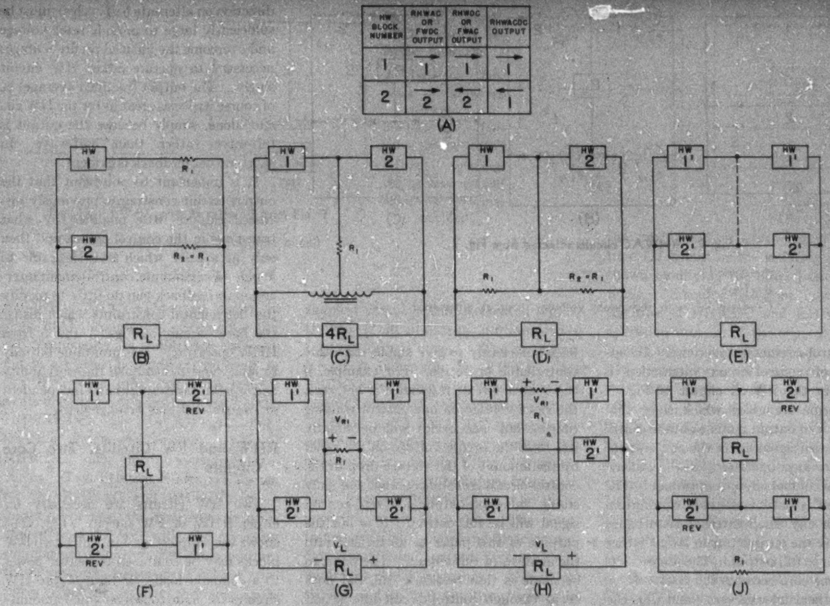


Fig. 6. RHW and FW output waveforms; two bias conditions shown in (A) through (C)

A—RHWAC; a-c component reverses, d-c does not
B—RHWDC; d-c component reverses, a-c does not
C—RHWACDC; a-c and d-c components reverse together
D—FWAC
E—FWDC



HW BLOCK NUMBER	RHWAC OR FWDC OUTPUT	RHWDC OR FWAC OUTPUT	RHWACDC OUTPUT
1	→ 1	→ 1	→ 1
2	→ 2	← 2	← 1

Fig. 7. RHW and FW block diagrams. REV under a block reverses arrows

A—Key
B—Type I, I′ for $R_1, R_2>0$
C—Type II, II′ for $R_1>0$
D—Type III
E—Type IV: Type I with HW block split, dotted lines join equipotential points
F—Type IV: (E) redrawn
G—Resistor added to type IV
H—Type IV″: (G) with a new twist
J—Type IV″: (H) redrawn

needs to be completely shut off. In all the circuits to be analyzed it will be necessary to keep careful track of both the prefire and reset constraint conditions.

Consider the circuit of Fig. 4 used for RHW output. This is a circuit in which the two HW circuits reset on the same half-cycle and fire on the same half-cycle. Here the two halves of the circuit do not interfere with one another's reset ($|V_r| \leq |e_s|$ as though the halves were completely independent), but they do interfere with one another's prefire voltage. The prefire voltage on either core is $|V_p| \geq |e_s| + |V_o|$. For this circuit, therefore, one of the HW circuits cannot be completely shut off while the other is fully firing, and full RHW operation is impossible.

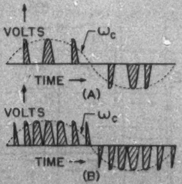


Fig. 8. RHWDC (A) and RFWDC (B) output waveforms, $\omega_c>0$

INTERACTIONS IN CONTROL CIRCUITS

Although consideration of control and bias circuitry is omitted in the sections to follow, the discussion at this point would not be complete without a brief mention

Table I. Reset and Prefire Constraints of RHW and FW Circuits

Circuit	Type of Output RHWAC or FWDC	 RHWDC or FWAC	 RHWACDC
I, II, IV	$\lvert V_r\rvert \leq \lvert e_s\rvert + \lvert V_0\rvert$	$\lvert V_r\rvert \leq \lvert e_s\rvert - \lvert V_0\rvert$	$\lvert V_r\rvert \leq \lvert e_s\rvert$
	$\lvert V_p\rvert \geq \lvert e_s\rvert$	$\lvert V_p\rvert \geq \lvert e_s\rvert$	$\lvert V_p\rvert \geq \lvert e_s\rvert + \lvert V_0\rvert$
I′, II′, IV′	$\lvert V_r\rvert \leq \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{(R_L+R_1)}$	$\lvert V_r\rvert \leq \lvert e_s\rvert - \frac{R_L\lvert V_0\rvert}{(R_L+R_1)}$	$\lvert V_r\rvert \leq \lvert e_s\rvert$
	$\lvert V_p\rvert \geq \lvert e_s\rvert$	$\lvert V_p\rvert \geq \lvert e_s\rvert$	$\lvert V_p\rvert \leq \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{(R_L+R_1)}$
II″, IV″	$\lvert V_r\rvert \leq \lvert e_s\rvert + \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert$	$\lvert V_r\rvert \leq \lvert e_s\rvert - \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert$	$\lvert V_r\rvert \leq \lvert e_s\rvert$
	$\lvert V_p\rvert \geq \lvert e_s\rvert$	$\lvert V_p\rvert \leq \lvert e_s\rvert$	$\lvert V_p\rvert \geq \lvert e_s\rvert + \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert$
III	$\lvert V_r\rvert \leq \lvert e_s\rvert - \frac{R_1\lvert V_0\rvert}{(R_L+R_1)}$	$\lvert V_r\rvert \leq \lvert e_s\rvert + \frac{R_1\lvert V_0\rvert}{(R_L+R_1)}$	$\lvert V_r\rvert \leq \lvert e_s\rvert$
	$\lvert V_p\rvert \geq \lvert e_s\rvert$	$\lvert V_p\rvert \geq \lvert e_s\rvert$	$\lvert V_p\rvert \geq \lvert e_s\rvert - \frac{R_1\lvert V_0\rvert}{(R_L+R_1)}$

Reset voltage $V_r = V_\theta$, $-\pi<\theta<0$
Prefire voltage $V_p = V_\theta$, $0<\theta<\theta_s$

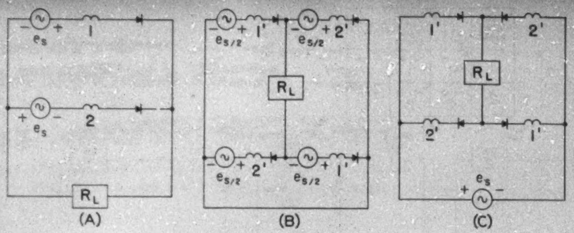


Fig. 9. RHWAC circuits selected from Fig. 7

A—Type I
B—Type IV
C—Type IV with redundant voltage, sources eliminated

of control-circuit interactions. An example of control-circuit interaction is shown in the FWAC circuit of Fig. 5. Waveforms are given which show that any value of output may be obtained with zero-control signal under the assumptions of square-loop operation, ideal rectifiers, and zero control-circuit impedance. The condition is one of unstable equilibrium in which any small external disturbance will cause the output to run away either to zero or to full output. The prefire voltage of one core becomes the reset voltage of the other and vice versa, and thus the circuit can maintain any value of output without a control signal. The circuit exhibits unity-positive feedback through control-circuit interaction.

The ideal circuit with unity-positive feedback just described is uncontrollable and useless as an amplifier. In actuality, however, slight deviations from the ideal conditions discussed will decrease the feedback sufficiently to give stable, smoothly controllable operation. For example, if the control-circuit resistance is not zero, the reset voltage on one control winding (during any one cycle) will be slightly less than the prefire voltage on the other by the amount of the voltage drop in the control-circuit resistance, and the only stable value of output for zero-control signal will be full output. It is not the purpose of this paper to go further into the analysis of this circuit. The point to be made is that feedback can and does occur through control-circuit interaction when the control windings are connected in certain ways.

If the control winding on core *2* is reversed, the positive feedback feature disappears. The prefire voltage induced in the control circuit opposes rather than aids the proper reset of the other core. The control voltage (which must now be a-c in order to reset both cores in the same direction on alternate half cycles), must be sufficiently large to provide reset voltage and overcome the induced prefire voltages necessary to operate either HW circuit singly. The output (rectified average) is, of course, twice as great as for the HW circuit alone, simply because the output is full-wave rather than half-wave. In short, no net feedback is evident.

It is important to point out that the output circuit constraints previously discussed can be little modified by what transpires in the control circuit and then only at a price which it is desirable to avoid. For example, control-circuit interaction or feedback can do little to modify the detrimental constraints which disbar the basic circuit of Figs. 3 and 5 from RHW operation. It is profitable initially to study and to establish the workability and worth of output circuits independent of control and bias arrangements.

RHW and FW Circuits, Two Core Circuits

Two HW circuits are necessary for either RHW or FW output. Fig. 7(B) shows the simplest way in which two HW blocks may be combined to deliver power to a common load. If both of the HW circuits are made to deliver equal amounts of output on alternate half-cycles, the result is FW output, either FWAC or FWDC. Waveforms are shown in Fig. 6(D) and (E). If one of the HW circuits is kept shut off while the other continuously delivers output and vice versa, the result is RHW output. There is more than one type of RHW output. There are, in fact, three: RHWAC, RHWDC

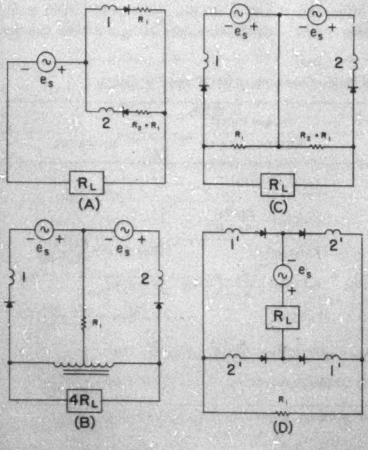
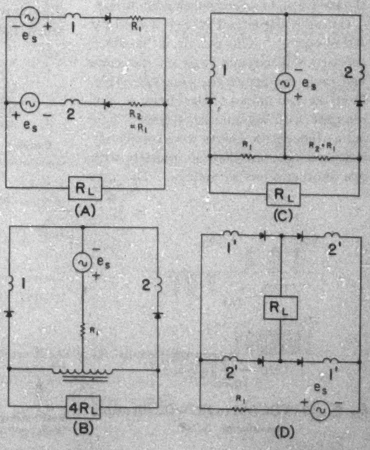


Fig. 10 (left). RHWDC circuits selected from Fig. 7

A—Type I′
B—Type II″
C—Type III
D—Type IV″

Fig. 11 (right). RHWACDC circuits selected from Fig. 7

A—Type I′
B—Type II″
C—Type III
D—Type IV″



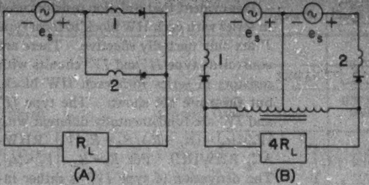


Fig. 12. FWAC circuits selected from Fig. 7

A—Type I; full-wave a-c or doubler circuit
B—Type II

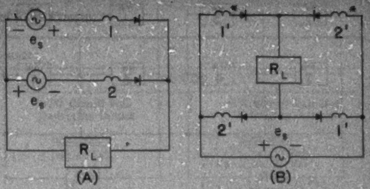


Fig. 13 (right). FWDC circuits selected from Fig. 7

A—Type I; full-wave or center-tap circuit
B—Type IV; full-wave bridge circuit
* These windings can be removed and remaining windings doubled

and RHWACDC. These are shown in Figs. 6(A) through (C). Strong a-c and d-c components are simultaneously present in all cases, a condition inherent in the HW waveforms, but both components do not reverse together except in the RHWACDC case. In RHWAC, only the a-c component reverses while in RHWDC only the d-c component reverses. Two bias conditions are shown for the RHWAC and RHWDC waveforms. For one, the nonreversing component goes to zero with the reversing component. For the other bias condition, the nonreversing component is kept relatively constant as the reversing component goes through zero. (Only the a-c fundamental goes through zero in the latter RHWAC case.) In the RHWACDC waveforms one bias condition is shown in which both cores fire in the same half-cycle to give an output that begins as a spike at midcycle and widens out into a half sine wave. An important point to notice is that in the RHWACDC case, the two HW circuits reset (or tend to reset) on the same half-cycle and fire (or tend to fire) on the same half-cycle. For the other RHW and FW circuits the HW circuits reset on alternate half-cycles.

BLOCK DIAGRAMS

Fig. 7 shows the various ways in which two HW circuits may be combined to deliver RHW or FW output to a common load. The key of Fig. 7(A) indicates the manner in which each of the block configurations can be used for the different RHW and FW modes of operation. The simplest circuit is that designated type *I*, the type *II* circuit uses an autotransformer (or transformer) output. The autotransformer is inoperable for RHWAC and FWDC because the d-c component in the output saturates the autotransformer. RHWDC and RHWACDC operation are possible provided the d-c control frequency ω_c is actually a low frequency, $0<\omega_c<\omega_s$ where ω_s is the supply frequency. (Fig. 8 shows RHWDC and RFWDC operating with $0<\omega_c<\omega_s$.) The autotransformer must be wound to accommodate maximum output voltage and minimum signal frequency. Mixing resistors replace the autotransformers in the type *III* circuit. Mixing resistors are inefficient, giving a maximum efficiency of only 17.1% with $R_L/R_1=\sqrt{2}$. The type *IV* bridge circuit is equivalent to the type *I* is as made evident by the method of derivation.

If the types *I*, *II*, and *IV* circuits operated without difficulty to deliver all RHW and FW types of output, there would be little necessity for considering the mixing-resistor arrangement of the type *III* circuit or the isolating resistors shown modifying the other basic circuit types. Difficulty does arise, however, in the case of RHWDC and RHWACDC operation because of the type of output circuit interactions discussed previously. These interactions are summarized in the reset and prefire constraint equations of Table I, where data are easily derived using the reasoning already developed. Specifically, it was previously established that RHW operation is impossible for a circuit in which $|V_p|\geqslant|e_s|$ and $|V_r|\leqslant|e|_s-|V_o|$. For $|V_p|\geqslant|e_s|$ and $|V_r|\leqslant|e|_s-k|V_o|$, $0<k<1$, limited RHW output is possible, but short of $k=0$, one HW circuit cannot be shut off completely while its companion is delivering full output. For $k>0$, $|V_r|$ is restricted to such a small value that $|\bar{V}_p|=|\bar{V}_r|$ can hold (as it must in steady state) only if the core which is supposed to be shut off saturates before the end of its fire half-cycle. Similarly, only limited RHW operation is possible in a circuit in which $|V_p|\geqslant|e_s|+|V_o|$ and $|V_r|\leqslant|e_s|$, another case discussed previously. The situation is improved with $|V_p|\geqslant|e_s|+k\ |V_o|\ 0<k<1$, but full RHW output is not possible short of $k=0$.

HW BLOCK NUMBER	RFWAC OUTPUT	RFWDC OUTPUT
1	$\overrightarrow{1}$	$\overrightarrow{1}$
2	$\overleftarrow{2}$	$\overrightarrow{2}$
3	$\overleftarrow{1}$	$\overleftarrow{1}$
4	$\overrightarrow{2}$	$\overleftarrow{2}$

Fig. 14 (left). Key to RFW block diagrams, Figs. 15–20. Blocks 1 and 2 deliver output of one phase (polarity); 3 and 4 deliver opposite phase (polarity)

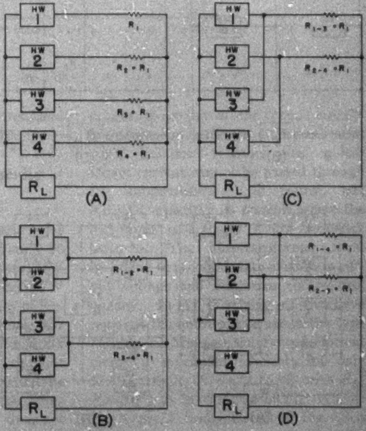


Fig. 15 (right). RFW type I block diagrams

A—Type I, I′ for R_1 to $R_4>0$
B—Type IA″
C—Type IB″
D—Type IC″

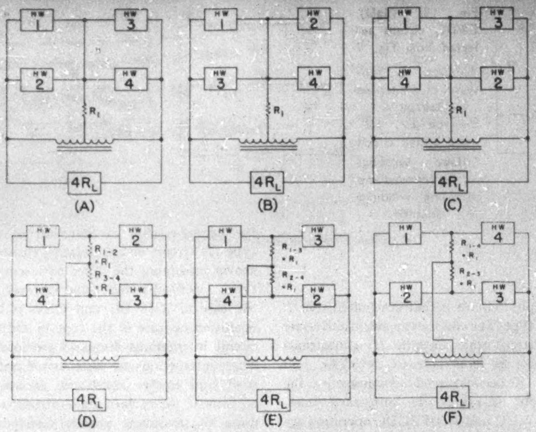


Fig. 16. RFW type II block diagrams; type II for isolating resistors absent

A–C—Types IIA‴ through IIC‴
D–F—Types IID″ through IIF″

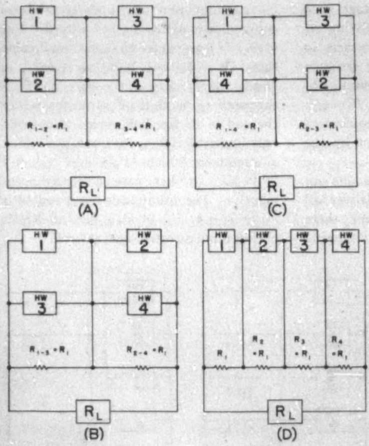
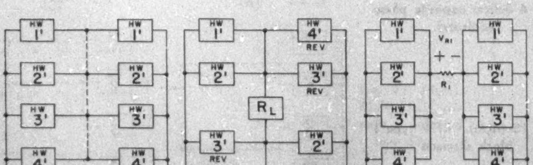


Fig. 17. RFW type III block diagrams

A–D—Types IIIA through IIID

Fig. 18 (below). RFW type IV block diagrams

A—Type IV: type I with HW blocks split; dotted lines join equipotential points
B—Type IVA redrawn
C—Resistors added



It is evident that isolation resistors used in series with each HW block to form type I' are only partially effective. There are equivalent type II' and IV' circuits with resistors in series with each HW block, but these are not shown. The type II'' and IV'' are fundamentally different with $|V_r| < |e_s| - (R_L - R_1)/R_L + R_1 |V_o|$ (RHWAC, RHWDC). For $R_1 = R_L$, $|V_r| < |e_s|$. The derivation of type IV'' is rather interesting. An attempt might be made to add a resistor in the type IV circuit as shown in Fig. 7(G), but there R_1 is of no benefit. With one HW circuit saturated, $V_L + V_{R1}$ is impressed across the other HW circuit exactly as though R_1 were immediately in series with R_L. In Fig. 7(H) the circuit is given a new twist and it can be seen that with one HW circuit saturated, $V_L - V_{R1}$ is impressed across the other HW circuit to create a new constraint situation. For $R_1 = R_L$, $V_L - V_{R1} = 0$, and the two HW circuits act independently. This is quite different from and much more beneficial than the action of the isolating resistors in the type I' circuit.

RHWAC CIRCUITS

Fig. 9 presents the detailed types I and IV output circuits selected from the block diagrams of Fig. 7. As reflected in Table I, these simplest and most efficient types offer no reset constraint problems and can deliver full RHWAC without difficulty. There is little cause, therefore, for using other types with isolating or mixing resistors. As has been mentioned previously, the type II circuit is inoperable for RHWAC because the d-c component in the output would saturate the autotransformer or transformer.

RHWDC CIRCUITS

Fig. 10 presents detailed RHWDC circuits selected from the several alternatives of Fig. 7. Table I reflects the disabling reset constraint of the simplest types I, II, and IV circuits used without isolating resistors. Of the circuits selected for Fig. 10, the type I' is relatively ineffective (this is one of the examples used in the section discussing reset constraints), and the type III is extremely inefficient (17.1% maximum efficiency). Types II'' and IV'' are the best choices. With $R_1 = R_L$, the interaction of the two component HW circuits is eliminated completely. The resulting circuit efficiency is 50%.

RHWADC CIRCUITS

Fig. 11 presents detailed RHWACDC output circuits selected from Fig. 7. The type I' circuit is relatively poor with

Table II. Reset and Prefire Constraints of RFW Circuits

Circuit	Type of Output			
	RFWAC		RFWDC	
	HW Circuit Pair Delivering Output	HW Circuit Pair Not Delivering Output	HW Circuit Pair Delivering Output	HW Circuit Pair Not Delivering Output
I, II, IV	$\lvert V_r\rvert \le \lvert e_s\rvert - \lvert V_0\rvert$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert + \lvert V_0\rvert \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \lvert V_0\rvert\end{array}\right\}$	$\lvert V_r\rvert \le \lvert e_s\rvert + \lvert V_0\rvert$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert - \lvert V_0\rvert \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \lvert V_0\rvert\end{array}\right\}$
I′, II′, IV′	$\lvert V_r\rvert \le \lvert e_s\rvert - \frac{R_L\lvert V_0\rvert}{R_L+R_1}$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1} \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}\end{array}\right\}$	$\lvert V_r\rvert \le \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert - \frac{R_L\lvert V_0\rvert}{R_L+R_1} \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}\end{array}\right\}$
IA″, IIA″, IVA″	$\lvert V_r\rvert \le \lvert e_s\rvert - \lvert V_0\rvert$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1} \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}\end{array}\right\}$	$\lvert V_r\rvert \le \lvert e_s\rvert + \lvert V_0\rvert$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert - \frac{R_L\lvert V_0\rvert}{R_L+R_1} \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}\end{array}\right\}$
IB″, IIB″, IVB″	$\lvert V_r\rvert \le \lvert e_s\rvert - \frac{R_L\lvert V_0\rvert}{R_L+R_1}$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1} \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \lvert V_0\rvert\end{array}\right\}$	$\lvert V_r\rvert \le \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert - \frac{R_L\lvert V_0\rvert}{R_L+R_1} \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \lvert V_0\rvert\end{array}\right\}$
IC″, IIC″, IVC″	$\lvert V_r\rvert \le \lvert e_s\rvert - \frac{R_L\lvert V_0\rvert}{R_L+R_1}$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert + \lvert V_0\rvert \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}\end{array}\right\}$	$\lvert V_r\rvert \le \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert - \lvert V_0\rvert \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}\end{array}\right\}$
IIA‴, IVA‴	$\lvert V_r\rvert \le \lvert e_s\rvert - \lvert V_0\rvert$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert + \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert\end{array}\right\}$	$\lvert V_r\rvert \le \lvert e_s\rvert + \lvert V_0\rvert$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert - \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert\end{array}\right\}$
IIB‴, IVB‴	$\lvert V_r\rvert \le \lvert e_s\rvert - \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert + \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \lvert V_0\rvert\end{array}\right\}$	$\lvert V_r\rvert \le \lvert e_s\rvert + \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert - \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \lvert V_0\rvert\end{array}\right\}$
IIC‴, IVC‴	$\lvert V_r\rvert \le \lvert e_s\rvert - \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert + \lvert V_0\rvert \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert\end{array}\right\}$	$\lvert V_r\rvert \le \lvert e_s\rvert + \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert - \lvert V_0\rvert \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{(R_L-R_1)}{(R_L-R_1)}\lvert V_0\rvert\end{array}\right\}$
IID″, IVD″	$\lvert V_r\rvert \le \lvert e_s\rvert - \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1} \\ \lvert V_p\rvert \le \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}\end{array}\right\}$	$\lvert V_r\rvert \le \lvert e_s\rvert + \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert - \frac{R_L\lvert V_0\rvert}{R_L+R_1} \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}\end{array}\right\}$
IIE″, IVE″	$\lvert V_r\rvert \le \lvert e_s\rvert - \frac{R_L\lvert V_0\rvert}{R_L+R_1}$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1} \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert\end{array}\right\}$	$\lvert V_r\rvert \le \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert - \frac{R_L\lvert V_0\rvert}{(R_L+R_1)} \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert\end{array}\right\}$
IIF″, IVF″	$\lvert V_r\rvert \le \lvert e_s\rvert - \frac{R_L\lvert V_0\rvert}{R_L+R_1}$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert + \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}\end{array}\right\}$	$\lvert V_r\rvert \le \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert - \frac{(R_L-R_1)}{(R_L+R_1)}\lvert V_0\rvert \\ \lvert V_p\rvert \ge \lvert e_s\rvert + \frac{R_L\lvert V_0\rvert}{R_L+R_1}\end{array}\right\}$
IIIA	$\lvert V_r\rvert \le \lvert e_s\rvert - \lvert V_0\rvert$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert - \frac{R_1\lvert V_0\rvert}{R_L+R_1} \\ \lvert V_p\rvert \ge \lvert e_s\rvert - \frac{R_1\lvert V_0\rvert}{R_L+R_1}\end{array}\right\}$	$\lvert V_r\rvert \le \lvert e_s\rvert + \lvert V_0\rvert$	$\left\{\begin{array}{l}\lvert V_r\rvert \le \lvert e_s\rvert + \frac{R_1\lvert V_0\rvert}{R_L+R_1} \\ \lvert V_p\rvert \ge \lvert e_s\rvert - \frac{R_1\lvert V_0\rvert}{R_L+R_1}\end{array}\right\}$

Constraints hold for normal operation in which only one HW circuit at a time is saturated. Reset voltage $V_r = V_g$, $-\pi<\theta<0$. Prefire voltage $V_p = V_g$, $0<\theta<\theta_s$. $\lvert V_p\rvert \ge \lvert e_s\rvert$ if not stated otherwise

full output impossible short of $R_1/R_L=\infty$. The type *II″* circuit delivers full output at 50% efficiency for $R_1=R_L$ provided the control-signal frequency is $0<\omega_c<\omega_s$. The type *III* circuit is, as always, inefficient. The type *IV″* circuit is equivalent to the type *II″* without the restriction on ω_c.

FW Circuits

There is no difficulty in obtaining either FWAC or FWDC from any of the types in Fig. 7 except for the impossibility of obtaining FWDC with the autotransformer output of type *II*. The simplest and most efficient circuits, types *I*, *II*, and *IV*, are all equivalent. These are shown detailed for FWAC in Fig. 12. Types *I* and *IV* for FWDC are shown in Fig. 13.

RFW Circuits, Four-Core Circuits

RFW Block Diagrams

Four HW circuits are necessary for RFW operation. Figs. 14 through 20 show various ways in which four HW blocks may be combined to deliver RFW power to a common load. The key of Fig. 14 indicates the manner in which the block configurations can be used for either RFWAC or RFWDC. The simplest circuit is type *I* of Fig. 15(A). The type *II* circuits of Fig. 16 use an autotransformer or transformer output. It is not difficult to see that all the type *II* circuits are equivalent although differences arise when isolation resistors are added. Direct current cannot be passed through the autotransformer, of course, but RFWDC operation is possible where the d-c input is actually a low-frequency $0<\omega_c<\omega_s$. The autotransformer must be wound to accommodate maximum output voltage and minimum control frequency. In Fig. 17 mixing resistors have replaced the autotransformer in the type *III* circuit. Mixing resistors are inefficient giving a maximum efficiency of only 17.1% with $R_L/R_1=\sqrt{2}$ [7.1% with $R_L/R_1=3\sqrt{2}$ as in the case of Fig. 17(D)[. The type *IV* bridge circuit of Fig. 18 is

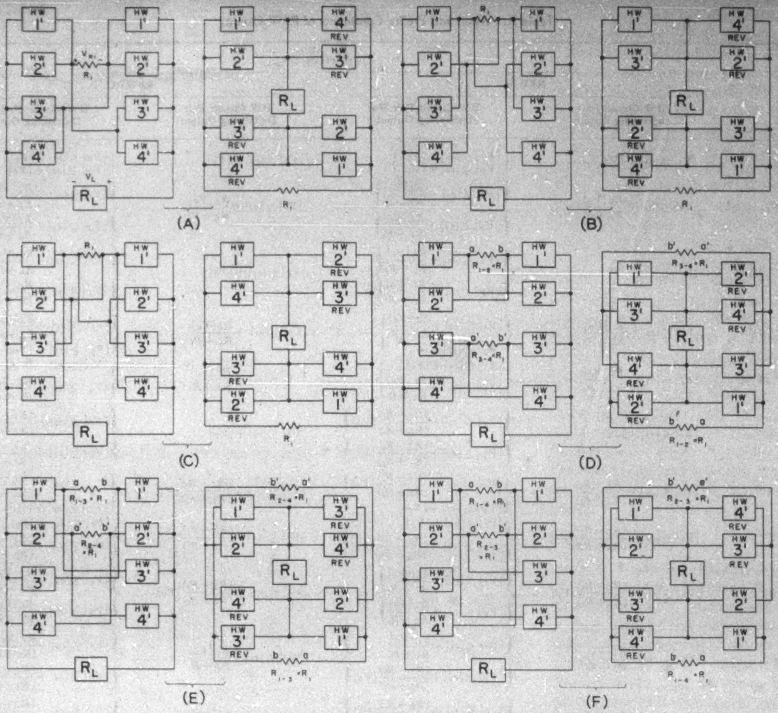


Fig. 19. RFW type IV'' and IV''' block diagrams, new twists to Fig. 18(C). Two drawings of same circuit in each case shown. In (D) through (F) one connection a–a', a–b', b–b', a'–b normally permissible

A–C—Types IVA''' through IVC'''1
D–F—Types IVD'' through IVF''

equivalent to the type *I* circuit as is made evident by the manner in which it is derived. The number of cores used in the bridge arrangement remains four since one core still suffices for a pair of windings when HW circuits are split in two as in Fig. 18(A).

If the basic types *I*, *II*, and *IV* circuits operated without difficulty to deliver both RFWAC and RFWDC, there would be little need to consider the addition of isolating resistors or the use of mixing resistors. Difficulty does arise, however. Types *I*, *II*, and *IV* circuits are completely inoperable for RFWDC. If the operation of the type *I* circuit is examined in detail following the reasoning developed earlier regarding reset and prefire interactions, it is not difficult to confirm the constraint conditions set forth in Table II. The constraints indicate that RFWAC operation offers no basic difficulty. The pair of cores delivering output (*1* and *2* for instance) have $|V_r| \leq |e_s| - |V_0|$ and $|V_p| \geq |e_s|$ which requires equal reset and equal output on successive half-cycles as in the type *I* FWAC circuit. The pair of cores not delivering output (*3* and *4*) have $|V_r| \leq |e_s| + |V_0|$ and $|V_p| \geq |e_s|$, and can thus be reset (using appropriate control circuitry) by an amount necessary to keep them shut off.

A completely different situation exists when the type *I* circuit or its equivalents are used for RFWDC output. When one pair of cores is attempting to deliver output, the other pair cannot be shut off. For the cores which should be shut off, $|V_r| \leq |e_s| - |V_0| = 0$ at full output. Thus the cores cannot be reset at all, and an intolerable situation exists in which pairs of cores are firing simultaneously and large circulating currents flow. No output of any kind is delivered to the load under these circumstances. What can be done to break up the unwanted interactions of the RFWDC type *I* circuit? In order to shut off one HW circuit pair when the other is delivering full output, it is necessary to make the maximum allowable $|\bar{V}_r|$ equal to or greater than the minimum mandatory $|\bar{V}_p|$. Then one of the HW circuit pairs can be shut off if the necessary reset voltage is provided in the control circuit. For the type *I* RFWDC circuit, the initial constraints on the pair of HW circuits which ought not to be delivering output are $|V_r| \leq |e_s| - |V_0|$ and $|V_p| \geq |e_s| + |V_0|$. Following the reasoning already developed, one way of allowing the pair to be shut off would be to make $|V_r| \leq |e_s|$, $|V_p| \geq |e_s|$ (if possible). Another equally good solution might be to make

Table III. Figure of Merit for RFWDC Circuits, Data from Table II

Circuit	HW Circuit Pair Not Delivering Output	
	$\|V_r\|/\|V_p\|$, $\|V_o\|=\|e_s\|$	$\|V_r\|/\|V_p\|$, $\|V_o\|=\|e_s\|$, $R_1=R_L$
I, IC″, II, IIC″, IIC‴, IV, IVC″, IVC‴	0	0
IB″, IIB″, IVB″	$R_1/2(R_L+R_1)$	0.25
I′, IA″, II′, IIA″, IID″, IV′, IVA″, IVD″	$R_1/(2R_L+R_1)$	0.33
IIB‴, IVB‴	$R_1/(R_L+R_1)$	0.50
IIE″, IVE″	$R_1/2R_L$	0.50
IIF″, IVF″	$2R_1/(2R_L+R_1)$	0.67
IIA‴, IVA‴	R_1/R_L	1.0
IIIA	$(R_L+2R_1)/R_L$	3.0

$|V_r| \leq |e_s|+|V_o|$, $|V_p| \geq |e_s|+|V_o|$. A partial solution can be obtained by placing a resistor in series with each HW circuit to form the type *I′* circuit. Limited but not full RFWDC output is then possible with $|V_r| \leq |e_s|-k|V_o|$, $|V_p| \geq |e_s|+k|V_o|$, $k=R_L/(R_L+R_1)$, as shown in Table II.

Various other ways of adding isolating resistors to the basic type *I* circuit are shown in Fig. 15. Figs. 16 through 19 show similar systematic modifications of varying merit for the basic circuit types. Types *I′*, *II′*, and *IV′* use four isolating resistors; *I″*, *II″*, and *IV″* use two; and *II‴* and *IV‴* use one. Types *II′*, and *IV′*, equivalent to *I′*, and *IIA″*, *IIB″*, *IIC″* and *IVA″*, *IVB″*, *IVC″*, equivalent to *IA″*, *IB″*, and *IC″*, are not shown. Many of these circuits appear scattered throughout the literature with scanty indication of their relative merit or equivalence to other circuits. The key to relative merit (and equivalence) is contained in the constraints listed in Table II. Once the restraints in the final column of the Table are established, it is simple enough to judge the relative effectiveness of the circuits for RFWDC operation. A figure of merit can be obtained by considering the ratio $|V_r|/|V_p|$ with $|V_o|=|e_s|$ (case of full output being delivered by the other HW circuit pair). This ratio is shown in Table III. It is not difficult to see that $|V_r|/|V_p|=0$ indicates a completely unworkable circuit. $0<|V|_r/|V_p|<1$ indicates a circuit from which limited but not full RFWDC is possible. $|V_r|/|V_p| \geq 1$ indicates that full RFWDC output is possible. Table III evaluates the RFWDC circuits for the case $R_L=R_1$. Note that the only circuits, other than the inefficient type *III* circuits, capable of giving full RFWDC output are the type *IIA‴* and *IVA‴* circuits.

The derivation of the type *IV* circuits is similar to that of the *IV* RHW circuits. In Fig. 18(C) an isolating resistor is added, but V_{R1} adds to V_L, and the effect is no different than that of putting R_1 directly in series with R_L. In Fig. 19(A), the reverse resistor connection causes V_{R1} to subtract from V_L. If $R_1=R_L$, $V_L-V_{R1}=0$, and all interaction between the two pairs of HW circuits is cancelled. Other possible isolationresistor combinations are explored in Fig. 19 and appraised in Tables II and III.

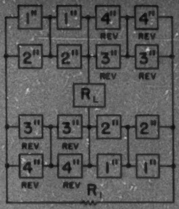


Fig. 20. RFW block diagram of bridge of bridges, equivalent to type IVA″

More complex forms of RFW circuits may be spun out indefinitely by splitting parallel pairs of HW circuits into bridge circuits in any of the preceding RFW block diagrams. One example is shown in Fig. 20 with the detailed circuit shown in Fig. 22(L). This circuit is sometimes seen in the literature despite its apparent complexity.

A word with regard to complexity is appropriate here. In all the RFW circuits shown so far and in the figures to follow, the minimum number of cores required is the same: four. Furthermore, the total volt-amperes ratings of the voltage sources, the windings, and the rectifiers (back voltage times forward current) is the same for all the circuits discounting whatever extra capacity must be provided for circulating currents and power dissipated in isolating and mixing resistors. Thus when the type *I* RHW or RFW circuit is split into a type *IV*

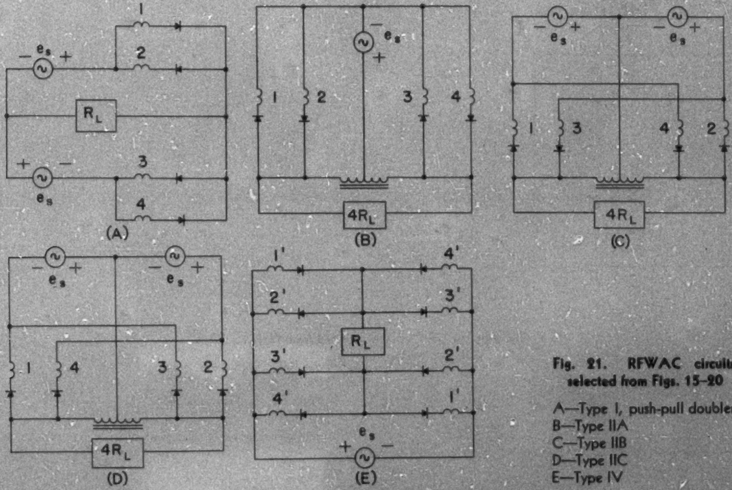


Fig. 21. RFWAC circuits selected from Figs. 15–20

A—Type I, push-pull doubler
B—Type IIA
C—Type IIB
D—Type IIC
E—Type IV

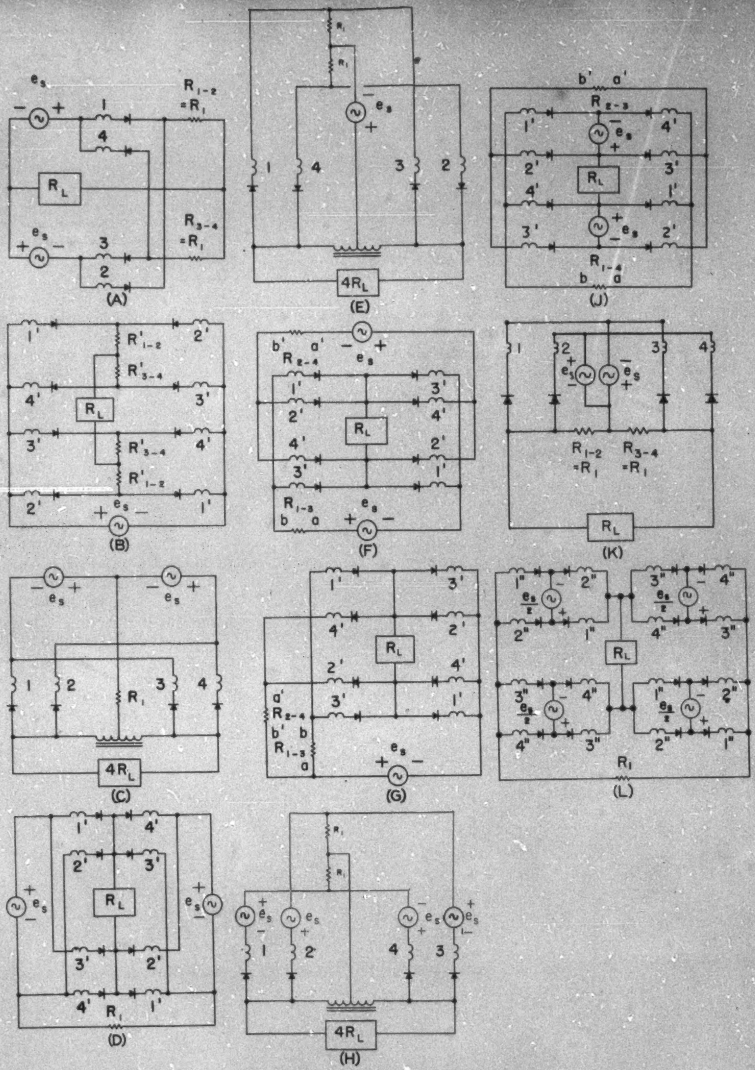

e_s
R_{1-2} = R_1
R_L
R_{3-4} = R_1
(A)
R_1
e_s
4R_L
(E)
b' a'
R_{2-3}
e_s
R_L
R_{1-4}
b a
(J)
R'_{1-2}
R'_{3-4}
R_L
R'_{3-4}
R'_{1-2}
e_s
(B)
b' a'
R_{2-4}
e_s
R_L
R_{1-3}
b a
e_s
(F)
R_{1-2} = R_1
R_{3-4} = R_1
R_L
(K)
e_s
R_1
4R_L
(C)
R_L
a'
R_{2-4}
b' b
R_{1-3}
a
e_s
(G)
e_s/2
R_L
R_1
(L)
e_s
R_L
R_1
(D)
R_1
e_s
4R_L
(H)

Fig. 22. RFWDC circuits selected from Figs. 15–20

A—Type IA″
B—Type IVA″, equivalent to IA″
C—Type IIA‴
D—Type IVA‴, equivalent to IIA‴
E—Type IIE″
F—Type IVE″, equivalent to IIE″
G—Type IVE″ with a—b′ connection
H—Type IIF″
J—Type IVF″, equivalent to IIF″
K—Type IIIA
L—Bridge of bridges, Fig. 20; pairs of windings may be removed as in Fig. 13(B)

bridge circuit, the number of load windings is doubled but the number of turns in each winding is cut in half, the number of rectifiers is doubled, but the back-voltage ratings on each rectifier may be cut in half, etc. The bridge arrangement obviates the need of input and/or output transformers (or autotransformers) in some cases. In any event, the simplest circuit on paper is not necessarily the most suitable circuit in practice.

RFWAC CIRCUITS

Fig. 21 shows several detailed RFWAC circuits selected from the earlier block diagrams with redundant voltage sources eliminated. In all cases cores *1* and *2* deliver output of one phase while *3* and *4* deliver output of the other phase. The basic types *I*, *II*, and *IV* circuits operate very well for RFWAC, and there is ordinarily no need to consider circuits with isolating or mixing resistors. Types *I*, *II*, and *IV* are all equivalent as indicated in Table II.

RFWDC CIRCUITS

Fig. 22 shows a wide selection of detailed RFWDC circuits from the earlier block diagrams. In all cases cores *1* and *2* deliver output of one polarity while *3* and *4* deliver output of the opposite polarity. Only the types *IIA'''*, *IVA'''*, the bridge of bridges, and the highly inefficient type *III* circuits can achieve full RFWDC output as indicated in Tables II and III. The other circuits are less successful but usable, and many appear in the literature.

Conclusions

This paper attempts to bring together a comprehensive collection of self-saturating magnetic-amplifier output circuits. A method of derivation is used which is intended to include as many known circuits as possible and to give a process of arriving at complex circuitry that may be remembered when details of the circuits themselves may be in doubt or forgotten. A simple analysis technique is used which makes troublesome interactions a little better defined and therefore a little more understandable. If the presentation has been successful, the paper should be helpful both in the selection of the most suitable circuits for particular jobs and in the better understanding of magnetic-amplifier circuitry and operation.

Reference Data

It is beyond the scope of this paper to give precise and detailed credits for first usage of the various circuits presented. Rather than give cursory credits, the author chooses to omit completely all such references. For details of this nature reference 1 is one of the best single sources. It discusses a large number of circuits, has a good bibliography, and is particularly thorough in its documentation of original publications and patents. Reference 2 is a short survey which includes several types of magnetic amplifiers other than the type dealt with in the present paper. For even broader perspective, reference 3 gives a brief review and bibliography of nonlinear magnetic devices, of which amplifiers are but one class. Reference 4 is an excellent source for detailed analysis of self-saturating and other magnetic amplifiers and has chapters and bibliography devoted to circuits and applications.

References

1. MAGNETIC-AMPLIFIER CIRCUITS (book), W. A. Geyger. McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1957.

2. TYPES OF MAGNETIC AMPLIFIERS—SURVEY, J. G. Miles. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 71, July 1952, pp. 229–38.

3. APPLICATIONS OF NON-LINEAR MAGNETICS, H. F. Storm. "Proceedings, Special Technical Conference on Magnetic Amplifiers," *AIEE Special Publication T98*, Aug. 1957, pp. 205–29.

4. MAGNETIC AMPLIFIERS (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955.

A New Dynamic Impedance Matching Circuit as Applied to a Conference Repeater

O. D. GRANDSTAFF
MEMBER AIEE

MOST telephone conversations are made between two parties. Occasionally the conversation may involve three parties, where the third party talks on an extension telephone. Another form of group conversation involves the interconnection of several telephone lines for a conference call. Here the lines are connected together in a conference circuit at the telephone exchange in a manner to provide optimum transmission and impedance matching. Small conference groups can be coupled through repeating coils but larger groups and conference circuits employed on toll switchboards need an amplifier to maintain satisfactory transmission levels. Current practice is to use telephone repeaters for this purpose.

During the establishment of a conference call, lines are added successively to the conference circuit until the desired number of parties are connected. Also, some parties may disconnect before the conference is concluded. These changes substantially alter both the impedance and voice signal power requirements in the conference circuit. Dummy loads are usually connected when the party disconnects to improve the stability but precise compensation cannot be obtained by this method as the impedance of lines connected to a conference circuit may vary considerably. Therefore, for the new conference repeater an amplifier with special input circuits was required that could keep each line properly terminated and operate with good stability regardless of the number of lines set up for a conference call. Since modern transmission standards provide uniform transmission levels in both local and toll switchboards, no gain is required in the conference re-

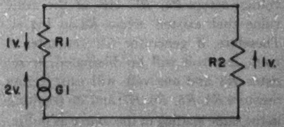


Fig. 1. Schematic of a telephone and load correctly impedance matched

---

Paper 59-824, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 20, 1959; made available for printing April 27, 1959.

O. D. GRANDSTAFF is with General Telephone Laboratories, Inc., Northlake, Ill.

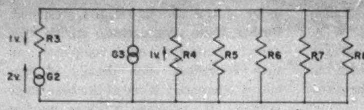


Fig. 2. Schematic of a telephone and conference circuit correctly impedance matched

peater but the original level must be maintained with any number of conference lines connected.

These requirements impose several problems in the conference repeater circuit. First, each line should be properly matched for impedance by the input of the conference repeater. A mismatch can adversely affect the antisidetone circuit of the telephone or the hybrid balance of a toll circuit repeater. Second, the amplifier must have the unique characteristic of delivering the same output level regardless of variations of load. A third requirement is that any party in the conference must be able to talk at will and all other parties hear him.

This conference repeater circuit is arranged to meet these requirements with as many as 12 lines. A new dynamic impedance matching circuit provides stable operation without need of a balancing network normally required in similar circuits where hybrids are used. The principles of operation of the conference repeater are as follows.

Fig. 1 shows the usual two party connection in a schematic diagram. Generator *G*1 represents the transmitter of the talking telephone. Resistor *R*1 is the internal impedance of the transmitting circuit. *R*2 is a load impedance representing the listening telephone. If the generator or talking telephone has an output of 2 volts, one volt is dissipated in the source *R*1 and one volt is dissipated in the load resistor *R*2. Under this condition the load impedance matches the source impedance and provides the ideal condition. Fig. 2 is a schematic of a conference repeater circuit where *G*2 represents the transmitter of the talking telephone; *R*3 its source impedance; and *R*4, *R*5, *R*6, *R*7, and *R*8 the load resistances of five listening telephones. These load resistors in parallel would normally cause a mismatch of impedance with the source impedance *R*3; but in this circuit a generator *G*3 has been added to produce a voltage across *R*4, *R*5, etc., of the same value that existed across *R*2 in Fig. 1. Therefore, if generator *G*2 generates 2 volts, one volt will be dissipated in resistor *R*3 and one volt will exist across resistors *R*4, *R*5, *R*6, *R*7, and *R*8 resulting in the same loading of the source as in Fig. 1. Under this condition the same current will flow from generator *G*2 of Fig. 2 as flows from generator of *G*1 Fig. 1. Also, each load resistor *R*4 to *R*8, will have one volt across it; the same as load resistor *R*2 of Fig. 1. In this manner the talking telephone is kept properly terminated and the several listening telephones receive full strength signal.

As long as generator *G*3 maintains the voltage across the load circuit, additional loads may be added or loads may be removed without altering the transmission level on the remaining circuits. In the conference repeater circuit, generator *G*3 is an electronic amplifier which uses an output circuit of the cathode follower type. It has sufficient inherent negative feedback that telephones may be added or removed with negligible change of its output voltage.

The diagram in Fig. 2 shows transmission only in one direction; from generator *G*2 and *R*3 to all the other loads. Since the requirements are that speech may originate from any telephone, the conference circuit must be a symetrical-type circuit arranged for any party to either talk or listen. This requirement presents a considerable problem since it involves 2-way voice transmission with an amplifier. Such an amplifier or voice repeater usually uses a hybrid circuit for detection of direction of power flow. In case of the conference repeater, hybrid circuits are not very practical because large impedance variations occur as several lines are switched onto or disconnected from the conference repeater. These variations cause excessive unbalance in the hybrid circuit which starts oscillation or singing in the circuit.

This conference repeater circuit uses a new method of detecting direction of power flow which does not require a balancing network. Line voltage and line currents are rectified by germanium diodes for comparison of their polarities to determine if the lines are carrying incoming or outgoing signals. Incoming currents control the amplifier but outgoing currents are prevented from having any effect on it. This is accomplished as shown in Fig. 3. Generator *G*4 and resistor *R*9 represent a talking telephone connected to line access circuit no. 1. Voice signals from *G*4 pass through transformer *TR*1, through capacitor *C*1, and resistor *R*11 to the main signal bus leads *A* and *B*. An amplifier, *AMPL*-1, has its cathode follower output connected to the *A* and *B* leads and functions the same as generator *G*3 of Fig. 2. The listening telephones such as *R*10 also are connected across the *A* and *B* leads. Load current flows from the *A* and *B* leads through resistor *R*12, capacitor *C*2, transformer *TR*2 to the load *R*10. A conference repeater may have several line access circuits like no. 1 and no. 2. All are alike and each may function at any time as a circuit for a talking or listening telephone. The signal voltage across leads *A* and *B* is applied through transformer *TR*5 to provide a voltage reference. During pos-

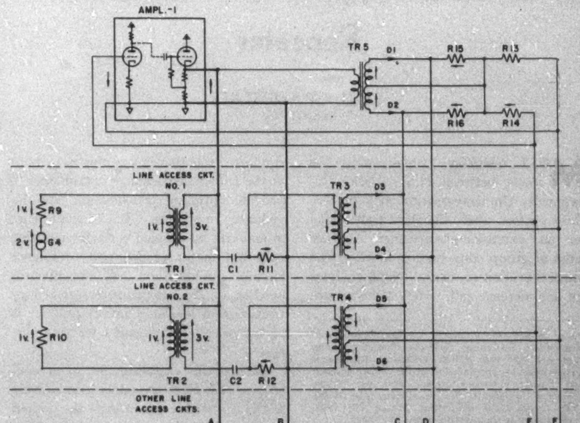


Fig. 3. Conference repeater circuit

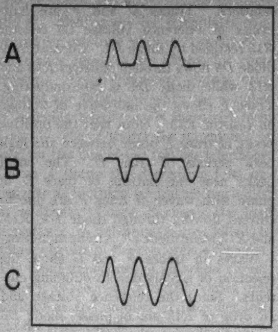


Fig. 4. Oscillogram of input signal

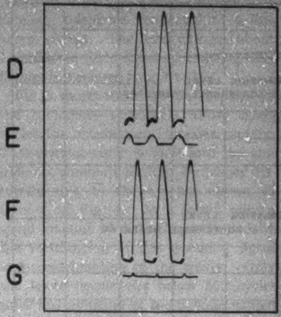


Fig. 5. Oscillogram of reference voltage and input signal

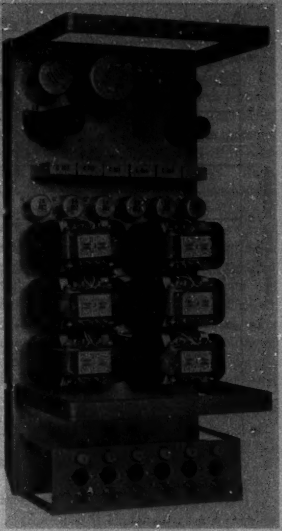


Fig. 6. Conference repeater for six lines

itive half-cycles of the *A* and *B* lead voltage, current flows through diode *D*1 and resistor *R*15 to the center tap of *TR*5. This causes positive half-cycles of the *A* and *B* lead voltage on resistor *R*15 while diode *D*2 prevents any reference voltage on *R*16.

The current references are obtained from resistors *R*11 and *R*12 which are small resistances approximately 5% of a line impedance connected to leads *A* and *B*. Signals flowing into or out of the *A* and *B* leads cause *IR* drops in these resistors of about 5% of the voltage across the *A* and *B* leads. The *IR* drop or voltage across *R*11 is proportional to the voice signal from generator *G*4. This small voltage is stepped up about 4 to 1 through transformer *TR*3 to have a better amplitude for diodes *D*3 and *D*4 and the input to the amplifier. On positive half-cycles of the *A* and *B* lead voltage, a secondary winding of *TR*3 causes current to flow through diode *D*3, resistor *R*16, and resistor *R*14, back to the winding of *TR*3. On negative half-cycles, current flows from the other winding of *TR*3 through diode *D*4, resistor *R*15, and resistor *R*13 back to the winding of *TR*3. The voltage drops developed across resistors *R*15 and *R*16 are of no consequence but the voltage drops across *R*13 and *R*14 are connected to the input of the amplifier. Resistor *R*12 has an *IR* drop proportional to the load current. The voltage drop is opposite in polarity to

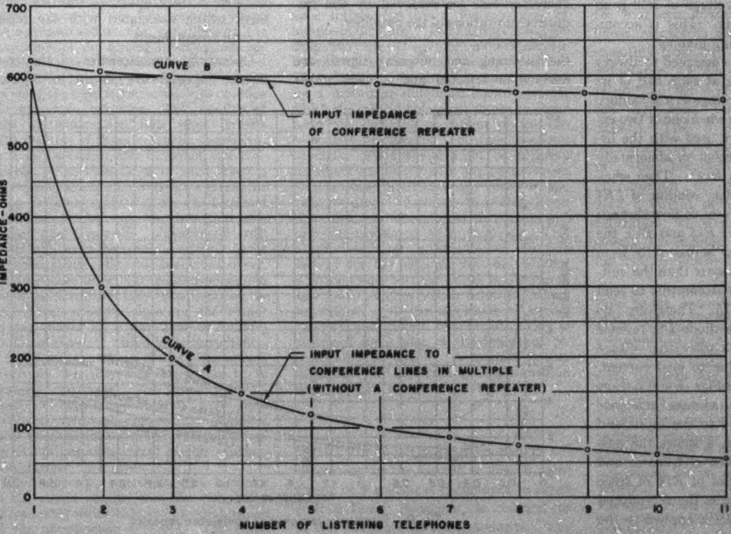


Fig. 7. Input impedance of conference repeater

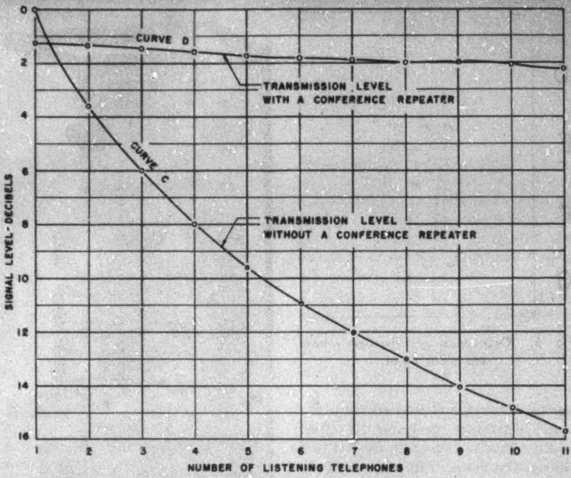


Fig. 8. Output level of conference repeater

that of the *IR* drop of *R*11, since it is due to current flowing out from the *A* and *B* leads. If this voltage were impressed on the input of the amplifier, it would subtract from the voltage applied to the amplifier from resistor *R*11, causing the input to be diminished and not be proportional to the incoming signal. Therefore, the circuit must function to prevent the outgoing signal voltage of *R*12 from reaching the amplifier. This is accomplished in the following manner.

Transformer *TR*5 is designed to always have a larger voltage at each half of its secondary than the voltage at a secondary winding of *TR*4 (even when one or two extension telephones are used with the listening telephone, causing an abnormally high voltage drop on *R*12). Then when voltage from a secondary winding of *TR*4 attempts to cause current to flow through diode *D*6 and resistors *R*15 and *R*13, the voltage drop already across *R*15 from transformer *TR*5 is greater than the voltage from *TR*4 that is attempting to send current through *R*15. Therefore, no current flows through diode *D*6, resistor *R*15, and resistor *R*13 when the voltage drop across *R*12 is due to load current.

The oscillograms in Figs. 4 and 5 show the relation of the vlotage reference waves and waves due to the input and output currents. Fig. 4 shows the positive half-cycles of *R*13 in trace *A* and the negative half-cycles of *R*14 in trace *B*. The sum of these is the reproduced signal of trace *C* which is applied to the input of the amplifier. Note the small notches, at the zero points of the cycles in trace *C*, caused by the addition of the rectified positive and negative half-cycles. These notches are kept to a minimum by use of gold-bonded germanium diodes for *D*3, *D*4, *D*5, and *D*6 and by circuit design with little phase shift. The notches are of high-frequency character and are removed by the frequency limitations of the amplifier.

The oscillogram in Fig. 5 shows how the incoming and outgoing signals are selected or rejected due to their phase relations. Traces *D* and *F* are half-cycles of reference voltage on resistor *R*15 rectified by diode *D*1. Current from diode *D*4 flows through resistors *R*15 and *R*13 while diode *D*1 is nonconducting. Trace *E* shows the half-wave of voltage on resistor *R*13. Note that the positive peaks of trace *E* occur between the high peak periods of trace *D*. The small peaks near the bottom of trace *D* in phase with waves of trace *E* are due to current flowing through diode *D*4, *R*15, and *R*13 while diode *D*1 is nonconducting.

Waves in traces *F* and *G* show the rejection of signals due to outgoing currents. Since the reference voltage and *IR* drop of *R*12 are in phase, diode *D*6 is conductive at the same time diode *D*1 causes large positive half-cycles on *R*15. The absence of positive half-cycles of the signal on trace *G* show how effectively signals due to outgoing currents are rejected. Only small notches remain which are lost in the amplifier like the notches in trace *C*.

A conference repeater equipped with six line-access circuits is shown in Fig. 6. The common amplifier is at the top of the unit. It is a conventional vacuum-tube amplifier using resistance-coupled triode voltage amplifiers to drive the cathode-follower output tube. One control, *R*6, sets the gain to provide the correct input impedance in the manner explained for generator *G*3 in Fig. 2. The access circuits are located in the middle of the unit. Equipment at the bottom includes a busy lamp, a test jack, and busy switch associated with the input of each access circuit.

Operating characteristics of the conference repeater are shown in Figs. 7

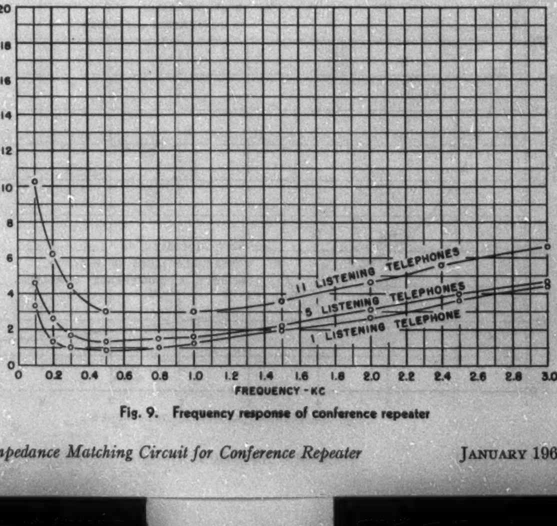


Fig. 9. Frequency response of conference repeater

and 8. Also curves for telephones connected in multiple without benefit of a conference repeater are shown for comparison. Voice frequency signals originate at one telephone while the number of listening telephones is increased from 1 to 11. Fig. 7 shows the impedance or load on the line of the talking telephone. Curve *A* shows the decreasing load impedance as listening telephones are added in multiple. Curve *B* shows the load impedance as telephones are added to access circuits of the conference repeater. The input impedance of an access circuit, while it is receiving a signal, varies only slightly as the number of listening telephones is increased.

Fig. 8 shows the signal level at the listening telephones in decibels. In curve *C* the level decreases rapidly as listening telephones are added in multiple. Curve *D* shows the level at a listening telephone connected to a line access circuit. The level changes about 1 decibel as the number of listening telephones is increased from 1 to 11.

Fig. 9 shows the frequency response of the conference repeater. Above 3,000 cycles the attenuation increases rapidly. This characteristic is built into the amplifier to eliminate the high-frequency distortion due to the notches shown in trace *C* of Fig. 4. As a result, only a small amount of distortion is caused by the rectification of the incoming signal and recombination of the waves. Only the lower frequencies below 500 cycles show any evidence of the notches caused by the recombination.

In a telephone conference, more than one person may speak into his telephone at the same time. Such coincidence of speech is more likely than where the group is assembled around a conference table. Nevertheless, the conference repeater can handle simultaneous speech; since each line-access circuit is always ready to receive incoming signals. To test simultaneous response, three persons were asked to read paragraphs from three unrelated subjects. Three other persons were instructed to listen and each to concentrate on his particular subject. Each one was able to understand completely the transmission of his subject.

The conference repeater provides equipment frequently needed for group telephone conversations. It can be used in toll and local offices or in private exchanges. It is easily applied to either manual- or automatic-type offices. Its constant-input impedance, fixed-gain amplifier, and inherent stability permit lines to be connected or disconnected without need of dummy loads or balancing networks.

Theory of Ferroresonance

JALAL T. SALIHI
ASSOCIATE MEMBER AIEE

Synopsis: The phenomenon of hysteresis occurring in a series nonlinear resonant circuit has interested engineers for a long time. The phenomenon is associated in literature with a resonance occurring at a critical value of applied voltage where the nonlinear inductance, variable with the magnitude of the forcing function, reaches a value so as to resonate with the fixed capacitor. This is probably the reason for using the term ferroresonance to describe the hysteresis behavior. The validity of such an explanation fails when circuits with reactors of sharp magnetic saturation, such as Deltamax and Orthonol, are considered. The inductance of such reactors may be approximated without appreciable error by linear regions of zero and infinite inductances. As will be shown in this paper, the hysteresis phenomenon still exists and even becomes more prominent as the magnetic saturation becomes sharper. On the other hand, the usual explanation given in literature fails when applied to such ideal circuits, since an inductance having values of either zero or infinity could not resonate with a finite capacitance.

Another explanation of the phenomenon given in the literature attributes it to negative resistance. The presence of regions of negative resistance have been confirmed through resort to graphical plots of equations derived. However, no direct mathematical proof of this explanation has yet been given.

The confusion concerning the behavior of the circuit results from the complexity of equations which hide the physical nature of the phenomenon involved. The equations derived in this paper are relatively simple and could be interpreted directly. This simplification is achieved by considering an idealized circuit consisting of a reactance with two regions of zero inductance interrupted by a linear region of infinite inductance. The analysis is further simplified by assuming a rectangular instead of a sinusoidal forcing function. The behavior of the circuit is thus expressed by a first-order difference equation, from which the critical points are shown to occur when the damping coefficient of the approximate differential equation becomes negative. Thus the concept of negative resistance associated with the phenomenon is confirmed mathematically without the need for graphical plots. It is also shown that the phenomenon of ferroresonance may be associated with feedback. This is expressed by a closed-loop diagram derived for the circuit. Another concept is introduced where the circuit is considered as a low-pass filter whose cutoff frequency depends on both the amplitude of forcing function and circuit parameters. This is expressed by a plot of the frequency response of the circuit. A method is also suggested for determining the transition time between the stable states.

THE HYSTERESIS phenomenon occurring in the circuit of Fig. 1, consisting of a nonlinear inductance in series with a linear capacitor, is referred to in the literature as ferroresonance. The phenomenon is explained as a resonance occurring at a critical value of the applied voltage where the inductance, variable with the applied voltage, reaches a value so as to resonate with the fixed capacitance at the frequency of the forcing function. Fig. 2 shows the ideal magnetization characteristic assumed. Figs. 3 and 4 show the correlation between the shape of magnetization characteristic and the hysteresis behavior. If the magnetic characteristic of the nonlinear reactor is assumed to be varied smoothly from one of gradually varying inductance to one of sharp magnetic saturation, the corresponding change in the hysteresis phenomenon will be smooth. The hysteresis phenomenon becomes more noticeable as the magnetic saturation becomes sharper. Thus the phenomenon of hysteresis associated with a circuit of very sharp magnetic saturation, which may be approximated closely by regions of infinite and zero inductance (curve *a*, Fig. 4), may be

---

Paper 59-788, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted July 28, 1958; made available for printing June 1, 1959.

Jalal T. Salihi, formerly with the Lenkurt Electric Company, Inc., San Carlos, Calif. is now in Iraq.

Portions of this paper have been abstracted from a thesis submitted to the graduate division of the University of California in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

The author wishes to acknowledge the suggestions and guidance of Dr. H. C. Bourne which made this work possible.

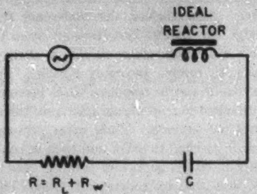


Fig. 1. Idealized basic ferroresonant circuit analyzed having an ideal reactor of zero winding resistance and the magnetization characteristic of Fig. 2

considered as a limiting case or an idealized ferroresonance. This is described by curve *a* of Fig. 3, such that, starting from zero, as the voltage is increased the reactor does not conduct until a critical voltage is reached at which the operation point suddenly jumps to P_1.

For any further increase in the voltage the current increases smoothly along the curve up to some point P_2. If the voltage is then reduced, the current reduces in the same manner to point P_1, where the voltage is again equal to its critical value. A further reduction in voltage is described by the path P_1P_3 along which the current reduces smoothly to point P_3 which corresponds to a second critical point at which a further reduction in voltage reduces the current abruptly to zero. The other two curves, *b* and *c* of Fig. 3, correspond to less ideal ferroresonance associated with the magnetization curves *b* and *c* of Fig. 4. As will be shown in this paper, the hysteresis phenomenon occurring in the ideal circuit is not a resonance, either instantaneous or average. This contradicts the explanation given in the literature, since such a circuit may be considered as a limiting case of less ideal ferroresonance circuits.

The basic ferroresonant circuit of Fig. 1 has been given considerable treatment due to both its practical and theoretical importance. The critical points at which the circuit becomes unstable were deduced either by graphical methods,[1] or from the knowledge of the behavior of one reactor by applying the principle of similitude,[2] or by fitting certain equations to measured relations.[3] In all these cases, a continuous magnetization characteristic of gradually varying permeability is assumed and a sinusoidal method of analysis is followed, where either the current or flux is assumed to be sinusoidal. The discrepancy between theoretical results obtained from a sinusoidal analysis and practical measurements increases as the saturation of the magnetic material of the core becomes sharper. The accuracy of the method may, of course, be increased by considering higher harmonics such as the third; however, consideration of harmonics higher than the third makes the sinusoidal approach to the analysis so tedious as to render it impractical.

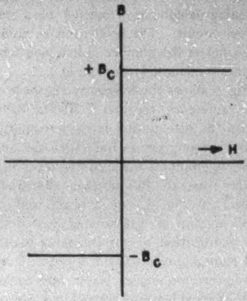


Fig. 2. Ideal magnetization characteristic assumed

In a previous paper by the author,[4] the steady-state behavior of an idealized ferroresonant circuit for the case of sinusoidal excitation was analyzed. The conduction angle was taken as an independent variable to which the various factors were related. The idealized equations were still so complicated that no direct mathematical explanation of the nature of the phenomenon could be seen. In this paper a rectangular excitation is assumed. This assumption simplifies the equation of the circuit appreciably so that the physical nature of the phenomenon hidden in previous analysis by the complexity of equations involved is exposed. Further, the dynamic equations of the circuit, from which useful conclusions may be made, are derived.

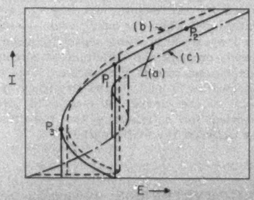


Fig. 3. Effect of the magnetization characteristic of the material used for the reactor of the series ferroresonant circuit on the steady-state behavior of the circuit

Method of Analysis and Assumptions

The idealized series resonant circuit of Fig. 1 consists of a nonlinear reactor in series with a fixed capacitance and resistance. The reactor is assumed to be ideal, having a magnetization characteristic composed of two linear regions of zero inductance interrupted by a linear region of infinite inductance as shown in Fig. 2. This circuit is referred to in this paper as the idealized ferroresonance circuit and the phenomenon of hysteresis occurring in this circuit as an idealized ferroresonance.

When an alternating forcing function is applied to the idealized ferroresonance circuit, the inductance of the circuit alternates between its zero and infinite value. During the intervals of infinite inductance, the circuit offers an infinite impedance to the flow of current. During the intervals of zero inductance, however, a current limited only by the capacitance and the resistance of the circuit flows. The angular interval through which current flows, referred to half a cycle of the forcing function, is called the conduction angle. The conduction angle is taken as an independent variable to which all the variables in the circuit are related. These variables are discontinuous functions of time. Thus, as a first step in the analysis, these variables are represented by artificial jump function. This function is of constant amplitude during each half-cycle of the forcing function and varies instantaneously to a new constant value at the beginning of the next half-cycle. Relating the variables thus expressed to the conduction angle leads to nonlinear difference equations. The nonlinear difference equations are approximated for small variations by linear difference equations. From these equations both the steady-state and the dynamic behavior of the circuit may be obtained.

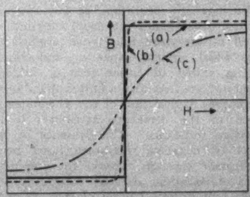


Fig. 4. Magnetization characteristic of the reactor of the series ferroresonant circuit corresponding to the steady-state behavior of Fig. 3

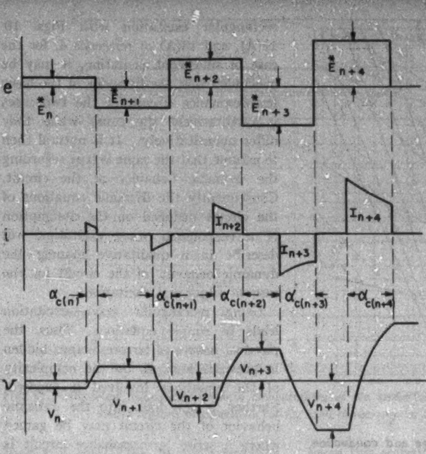


Fig. 5. Waveshape of current flowing in the idealized basic ferroresonant circuit of Fig. 1 and the voltage across the capacitor for a rectangular excitation voltage of varying amplitude applied to the circuit

The qualitative correlation of the characteristics derived in reference 4 for the case of sinusoidal excitation with the corresponding characteristics derived in this paper for the case of rectangular excitation gives an indirect proof of the validity of the analysis presented.

Modes of Operation and Circuit Equations

Let a rectangular voltage, of varying amplitude, be applied to the circuit of Fig. 1. The resulting current pulses and the voltage across the capacitor are given in Fig. 5, where α_c, V, i, and I denote the conduction angle, the constant voltage across the capacitor during nonconduction intervals, the instantaneous value of current, and the half-cyclic average value of the current, respectively. The subscripts n, $n+1$, $n+2$, ..., are used to assign these values to the corresponding nth, $n+1$, $n+2$, ..., half-cycle of the forcing function. During the conduction intervals, α_c, of each half-cycle of the applied voltage, the flux in the reactor is constant at its saturation value. Consequently, the induced voltage is zero and the reactor behaves as a resistance equal to its winding resistance. During the nonconduction intervals the flux in the reactor is changing. Consequently, a voltage will be induced across the reactor which behaves as an infinite inductance preventing the flow of current.

It is convenient to describe the behavior of the circuit in terms of three unknowns α_c, V, and I. Thus three difference equations are required to determine any of these unknowns. These equations may be derived as follows.

Equation 1

If Q_n and Q_{n+1}, are the charges across the capacitor during the nonconduction interval of nth and $n+1$ half-cycle, then:

$$Q_{n+1}=-Q_n+\frac{1}{\omega}\int_{\pi-\alpha n}^{\pi} i_n d\omega t=-Q_n+\frac{\pi}{\omega}I_n$$

since

$$Q_n=CV_n$$

and

$$\phi=\tan^{-1}(1/\omega CR)$$

then

$$\pi \tan \phi R I_n=V_n+V_{n+1} \qquad (1)$$

Equation 2

The differential equation of the circuit during the nth conduction interval is of the form:

$$Ri_n+\frac{1}{C}\int i_n dt=E^*$$

where E^* represents the amplitude of the rectangular excitation voltage. Solution of this equation is of the form:

$$i_n=K\epsilon^{-\frac{t}{CR}}=K\epsilon^{-\frac{1}{\omega CR}\omega t}=K\epsilon^{-\tan\phi\,\omega t}$$

Substituting the boundary condition

$$i_n=\frac{E_n^*+V_n}{R}$$

at

$$\omega t=\pi+\alpha_{c(n)}$$

the value of the constant K may be determined. The equation of current then becomes:

$$i_n=\frac{E_n^*+V_n}{R}\epsilon^{\tan\phi\,[\pi-\alpha_{c(n)}]}\times\epsilon^{-\tan\phi\,\omega t}$$

The half-cyclic average value I_n of the current may be obtained by integrating this equation between the limits

$$\omega t=\pi-\alpha_{c(n)}$$

to $\omega t=\pi$ which gives the second equation:

$$1-\epsilon^{-\tan\phi\,\alpha_{c(n)}}=\frac{\pi\tan\phi}{E_n^*+V_n}RI_n \qquad (2)$$

Equation 3

The differential equation of the circuit during the nth nonconduction interval is of the form:

$$N\frac{d\Phi}{dt}\times 10^{-8}=E_n^*+V_n$$

where

Φ = flux
N = number of turns on the core

Since $d\Phi=Ad\beta$, where A is the cross-sectional area of the core, one may write:

$$d\beta=\frac{E_n^*+V_n}{AN\omega\times 10^{-8}}d\omega t$$

Integrating this equation gives

$$\beta=K_1+\frac{10^8(E_n^*+V_n)}{AN\omega}\int_{\omega t_i}^{\omega t}d\omega t$$

At $\omega t=0$, $\beta=-B_c$ due to negative saturation, while at $\omega t=\pi-\alpha_{c(n)}$, $\beta=+B_c$, due to positive saturation. Inserting these conditions in the foregoing equation gives the third of the required equations, which is of the form:

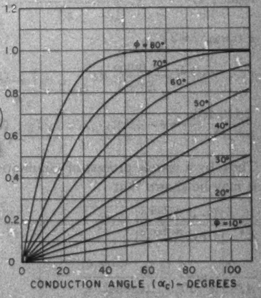


Fig. 6. Steady-state calculated relationship between conduction angle and constant voltage across the capacitor during presaturation intervals expressed as a fraction of the amplitude of the rectangular voltage applied to the basic ferroresonant circuit

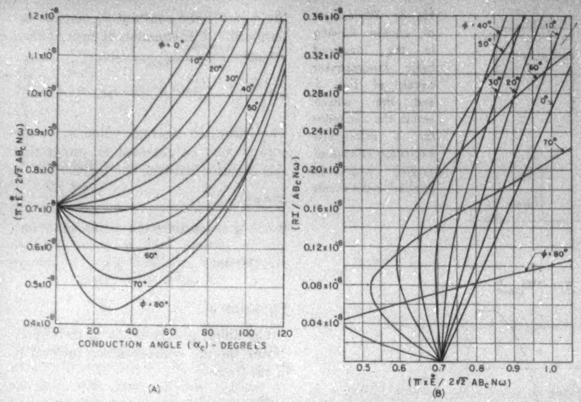


Fig. 7. Calculated steady-state relationships between (A) applied voltage and conduction angle and (B) applied voltage and half-cyclic average value of current, for a rectangular voltage applied to the basic ferroresonant circuit

$$\frac{2AB_cN\omega}{10^8(E_n^*+V_n)}=\pi-\alpha_{c(n)} \tag{3}$$

These three equations containing three unknowns give enough information to solve for both the steady-state and the transient behavior.

Steady-State Behavior

In steady-state operation

$$I_n=I_{n+1}=I_{n+2}\ldots$$

$$V_n=V_{n+1}=V_{n+2}\ldots$$

$$\alpha_{c(n)}=\alpha_{c(n+1)}=\alpha_{c(n+2)}$$

Thus the three equations derived reduce to

$$\pi \tan\phi RI=2V \tag{1A}$$

$$1-\epsilon^{-\tan\phi\,\alpha_c}=\frac{\pi\tan\phi}{E^*+V}RI \tag{2A}$$

$$\frac{2AB_cAN\omega}{10^8(E^*+V)}=\pi-\alpha_c \tag{3A}$$

Thus, by a simple process of elimination, it may be shown that

$$\frac{V}{E^*}=\frac{[1-\epsilon^{-\alpha_c\tan\phi}]}{[1+\epsilon^{-\alpha_c\tan\phi}]} \tag{4}$$

$$\frac{E^*}{AB_cN\omega}=\frac{1}{10^8}\,\frac{[1+\epsilon^{-\alpha_c\tan\phi}]}{[\pi-\alpha_c]} \tag{5}$$

$$\frac{RI}{AB_cN\omega}=\frac{2}{10^8}\,\frac{[1-\epsilon^{-\alpha_c\tan\phi}]}{\pi\tan\phi[\pi-\alpha_c]} \tag{6}$$

where $V/AB_cN\omega$, $E^*/AB_cN\omega$, and $RI/AB_cN\omega$ are the nondimensional values of voltage across capacitor, applied voltage, and the average value of current. Equation 4 is obtained by eliminating *RI* from equations 1(A) and 2(A), while equation 5 is obtained by eliminating *V* from equations 4 and 3(A). Equation 6 is obtained by eliminating E^*+V from equations 2(A) and 3(A). Fixing values of ϕ at 10, 20, 30, 40, 50, 60, 70, and 80 degrees, the nondimensional graphical relationship of Figs. 6 and 7 are obtained. Fig. 6 relates the voltage across the capacitor to the conduction angle as deduced from equation 4. Fig. 7(A) relates the applied voltage to the conduction angle as deduced from equation 5. Fig. 7(B) relates the voltage applied to the circuit to the average value of the current flowing. This is deduced from equations 5 and 6, where the corresponding values of voltage and current for any conduction angle may be calculated.

From a comparison of Figs. 6, 7(A), and 7(B) of this paper for the case of rectangular excitation with Figs. 10 11(A), and 13(A) of reference 4, for the case of sinusoidal excitation, it may be concluded that the behavior of the basic ferroresonance circuit for the two cases is qualitatively the same while they differ quantitatively. It is natural then to assume that the same is true regarding the dynamic behavior of the circuit. Consequently the dynamic equations of the circuit deduced on the assumption of a rectangular wave excitation will describe in a qualitative manner the dynamic behavior of the circuit for the case of sinusoidal excitation.

Using rectangular wave excitation leads to simple equations. Thus, the physical nature of ferroresonance hidden in previous analysis by the complexity of the equations involved is exposed. Further, a new look into the dynamic behavior of the circuit may be gained where a series ferroresonance circuit is considered as a low-pass filter whose frequency-response characteristic may be calculated from simple equations. From these characteristics the speed of response may be deduced.

Dynamic Behavior and Condition for Stability

The three difference equations involve nonlinear terms. These equations may be linearized for small variations about given operating points. To do this, equation 2 may be written as

$$1-\frac{\epsilon^{\tan\phi[\lambda-\alpha_{c(n)}]}}{\epsilon^{\tan\phi\,\lambda}}=\frac{\pi\tan\phi}{E_n^*+V_n}\,RI_n$$

where λ is a steady-state operation conduction angle about which the transient variation takes place. For small transient changes $(\lambda-\alpha_{c(n)})$ is small, and foregoing equation may be written as

$$1-\frac{1+\tan\phi\lambda-\tan\phi\,\alpha_{c(n)}}{\epsilon^{\tan\phi\,\lambda}}=\frac{\pi\tan\phi}{E_n^*+V_n}\,RI_n \tag{7}$$

By the process of substitution in equa-

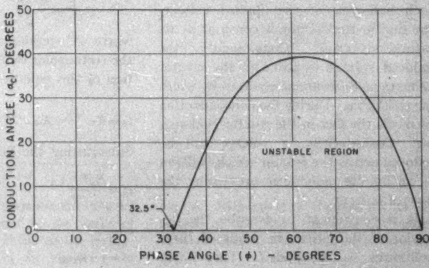


Fig. 8. Calculated critical value of conduction angle below which the basic ferroresonant circuit becomes unstable

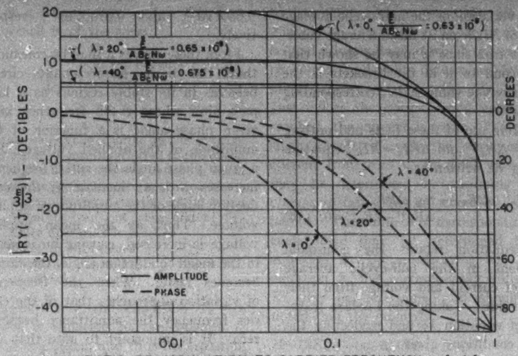


Fig. 9. Calculated frequency response characteristic of the basic ferroresonant circuit in the stable regions of operation for the case of rectangular excitation

tions 1, 3, and 7, the difference equation of the circuit is derived, which is of the form:

$$[K_3 - \pi \tan\phi]RI_n + K_3[RI_{n+1}] - [E_n{}^* + E_{n+1}{}^*] + \frac{4AB_cN\omega}{10^8}K_2 = 0 \quad (8)$$

where K_2 and K_3 are constants corresponding to a given steady-state operation conduction angle λ, and are given by

$$K_2 = \frac{\tan\phi}{\epsilon^{\lambda \tan\phi} + \tan\phi(\pi - \lambda) - 1}$$

$$K_3 = \frac{\pi \tan\phi\, \epsilon^{\lambda \tan\phi}}{\epsilon^{\lambda \tan\phi} + \tan\phi(\pi - \lambda) - 1} \quad (9)$$

Letting

$$I_{n+1} = I_n + \Delta I_n$$

and

$$E_{n+1}{}^* = E_n{}^* + \Delta E_n{}^*$$

equation 9 gives

$$K_3 \Delta RI_n + [2K_3 - \pi \tan\phi]RI_n - [\Delta E_n{}^* + 2E_n{}^*] + 4 \times 10^{-8} AB_cN\omega K_2 = 0 \quad (10)$$

for a constant applied voltage $\Delta E_n{}^* = 0$, and $E_n{}^* = E^*$. Thus equation 10 may be written

$$\Delta RI_n + \left[\frac{2K_3 - \pi \tan\phi}{K_3}\right]RI_n + \frac{2}{K_3}\left[\frac{2AB_cN\omega}{10^8}K_2 - E^*\right] = 0 \quad (11)$$

For slow variations this equation may be approximated by a differential equation of the form:

$$\frac{d}{dt}RI + \left[\frac{2K_3 - \pi \tan\phi}{K_3}\right]RI + \frac{2}{K_3}\left[\frac{2AB_cN\omega}{10^8}K_2 - E^*\right] = 0 \quad (12)$$

From this equation it may be concluded that the system is stable when the damping coefficient is positive. The condition for stability may be expressed as

$$\frac{2K_3 - \pi \tan\phi}{K_3} \geq 0 \quad (13)$$

Substituting for values of K_3 given by equation 9 in this equation the condition for stability becomes

$$[\epsilon^{\lambda \tan\phi} + \lambda \tan\phi - \pi \tan\phi + 1] \geq 0$$

At the boundary of the stable region $\lambda = \lambda_c$, thus

$$[\epsilon^{\lambda_c \tan\phi} + \lambda_c \tan\phi - \pi \tan\phi + 1] = 0 \quad (14)$$

From this equation the critical value of steady-state conduction angle corresponding to any given value of phase angle may be determined. The corresponding critical value of applied voltage and current may be calculated by substituting the critical value of λ_c, instead of α_c, in the steady-state equations 5 and 6, respectively. Values of λ_c corresponding to different values of ϕ are plotted in Fig. 8. The calculated curve is the boundary between stable and unstable modes of operation. For values of circuit parameters inside the region bounded by this curve and the X-axis, the circuit exhibits a hysteresis effect, while outside this region the behavior of the circuit is described by a single-valued function and the circuit is stable. From Fig. 8, it may be seen that no hysteresis effect exists if the phase angle is less than 32.5 degrees. The critical value of the phase angle $\phi = \tan^{-1} 2/\pi = 32.5$ degrees may be obtained by letting $\lambda_c = 0$ in equation 14.

In spite of the approximation applied, the condition of stability given by equations 13 or 14 is an exact condition because at the point where the circuit becomes just unstable, the subsequent transient starts with a slow change. During this instant, the approximate equation is exact. This may be verified graphically from a comparison of Fig. 8 with the minimum points of Fig. 7(A), deduced on the basis of exact steady-state equations showing that the results obtained from these two figures for the condition of stability are identical. An analytical proof of this may also be obtained from the investigation of the stability of the difference equation. The solution of this equation is of the form:

$$RI_n = D\left[1 - \frac{2K_3 - \pi \tan\phi}{K_3}\right]^n$$

where D is a constant corresponding to the initial conditions. From this solution it may be concluded that the circuit is unstable when the absolute value of the quantity in the bracket exceeds unity. This is true since n takes consecutive integer values, thus causing RI_n to increase indefinitely. This condition is, of course, limited by saturation effects in physical systems. The condition for stability of the difference equation may thus be expressed as

$$\left|1 - \frac{2K_3 - \pi \tan\phi}{K_3}\right| \leq 1$$

At the boundary of the stable region this equation may be written as

$$\left|1 - \frac{2K_3 - \pi \tan\phi}{K_3}\right| = 1$$

From this equation it may be concluded that the stable region has two boundaries defined by the following two equations:

$$+\left[1 - \frac{2K_3 - \pi \tan\phi}{K_3}\right] = 1$$

or

$$\frac{2K_3 - \pi \tan\phi}{K_3} = 0$$

$$-\left[1 - \frac{2K_3 - \pi \tan\phi}{K_3}\right] = 1$$

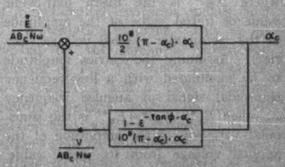

$\frac{E^*}{AB_cN\omega}$
+
$\frac{10^8}{2}(\pi - \alpha_c) \cdot \alpha_c$
α_c
$\frac{1 - \epsilon^{-\tan\phi \cdot \alpha_c}}{10^8(\pi - \alpha_c) \cdot \alpha_c}$
$\frac{V}{AB_cN\omega}$

Fig. 10. Equivalent feedback diagram for the basic ferroresonant circuit of Fig. 1 for the case of rectangular excitation

or

$\phi=0$

The boundary of stability defined by the first equation is the same as equation 13, derived for the differential equation, while the second equation indicates that the system becomes unstable at negative values of phase angle $\phi=\tan^{-1}(1/\omega cR)$, that is, at negative values of frequency which are physically unrealizable. From these results the following conclusions may be made concerning the behavior of the basic series ferroresonant circuit:

1. The circuit is stable for $\phi<\tan^{-1} 2/\pi=32.5$ degrees.
2. For values of $\phi>\tan^{-1} 2/\pi$ the circuit is unstable in some regions of its operation where the output changes in an exponential manner limited by saturation effects. Since the equation involved is of the first order, there is no possibility of sustained oscillation in this ideal circuit.
3. Unstable operation associated with negative phase angles predicted mathematically is physically unrealizable.

Frequency Response

The frequency response of the basic ferroresonant circuit in the stable regions may be obtained by assuming the amplitude of the rectangular applied voltage to be modulated with a low-frequency signal compared with the carrier. Consequently, the resultant variation in the half-cyclic average value of current will be sinusoidal also. This is true for small variations where the behavior of the circuit is described by linear equations. The speed of response may thus be obtained from the frequency-response characteristics expressing the ratio of the output current to the input signal at different signal frequencies, modulating the constant carrier frequency.

The dynamic behavior is given by equation 10 which may be rewritten as

$$K_3\Delta RI_n+[2K_3-\pi\tan\phi]\times\left[RI_n-\frac{2E_{DC}{}^*-4\times10^{-8}\times AB_cN\omega K_2}{2K_3-\pi\tan\phi}\right]-\Delta E_n{}^*-2[E_n{}^*-E_{DC}{}^*]=0 \quad (15)$$

where $E_{DC}{}^*$ is the mean half-cyclic value or initial value of the amplitude of the rectangular applied voltage (carrier). Assuming the carrier amplitude to be modulated with a low-frequency sinusoidal signal of angular frequency ω_m, then $E_n{}^*-E_{DC}{}^*=E_n{}^{*\prime}$ is the sinusoidal variation of the amplitude of the applied voltage from its mean value. Also

$$\left[RI_n-\frac{2E_{DC}-4\times10^{-8}\times AB_cN\omega K_2}{2K_3-\pi\tan\phi}\right]=RI_n{}'$$

is the variation of the half-cyclic average value of RI_n from its mean average value. This is true since it may be shown that the second term in the brackets is the mean average value of RI_n corresponding to $E_{DC}{}^*$.

Making use of these facts and putting $\Delta E_n{}^*=\Delta E_n{}^{*\prime}$ and $\Delta RI_n=RI_n{}'$, equation 16 may be written as

$$K_3\Delta RI_n{}'+[2K_3-\pi\tan\phi]RI_n{}'-\Delta E_n{}^{*\prime}-2E_n{}^{*\prime}=0 \quad (16)$$

This equation relates the sinusoidal variations in the half-cyclic average value of current and voltage from their mean value. Taking the Laplace transform of equation 16 and neglecting the initial conditions gives

$$K_3[\epsilon^{\frac{\pi}{\omega}s}-1]RI'(s)+[2K_3-\pi\tan\phi]RI'(s)-[\epsilon^{\frac{\pi}{\omega}s}-1]E^{*\prime}(s)-2E^{*\prime}(s)=0$$

This equation may be written as

$$\frac{E^{*\prime}(s)}{RI'(s)}=\frac{Z(s)}{R}=\frac{K_3+[K_3-\pi\tan\phi]\epsilon^{-\frac{\pi}{\omega}s}}{\left[1+\epsilon^{-\frac{\pi}{\omega}s}\right]}$$

where ω is the angular frequency of the applied voltage (carrier), and $\epsilon^{-(\pi/\omega)s}$, represents a time shift of half a cycle. The sinusoidal response of the circuit may be obtained by letting $S=j\omega_m$ where ω_m is the angular frequency of the modulating signal. Thus

$$\frac{Z(j\omega_m)}{R}=\frac{K_3+[K_3-\pi\tan\phi]\epsilon^{-j\pi\frac{\omega m}{\omega}}}{\left[1+\epsilon^{-j\pi\frac{\omega m}{\omega}}\right]} \quad (17)$$

where K_3 is given by equation 9.

The average admittance of the circuit to the low-frequency modulating signal $Y(j\omega_m)=1/Z(j\omega_m)$ may be obtained from this equation. Thus at $\phi=0$, $K_3=1$ and $Z(j\omega_m)/R=1$, showing that the average impedance $Z(j\omega_m)$ is equal to R, regardless of the frequency of signal or the mean conduction angle (function of applied voltage as given by equation 15). However, this is not true for the case of capacitive loading where the impedance becomes a function of both frequency and amplitude of forcing function. As an example, for $\phi=30$ degrees the value of the nondimensional admittance $RY(j\omega_m)$ is calculated and the frequency-response characteristics are shown in Fig. 9, which is a plot of the magnitude and phase of $RY(j\omega_m)$ for three different values of mean conduction angle (λ) of 0, 20, and 40 degrees taken as parameters. The corresponding three values of applied voltage may be obtained from Fig. 7(A), and are 0.63×10^{-8}, 0.65×10^{-8}, and 0.675×10^{-8}, respectively.

From these curves it may be concluded that the basic ferroresonant circuit behaves in its stable regions as a low-pass filter whose cutoff frequency at a given phase angle is a function of the amplitude of the applied voltage. For a given phase angle the cutoff frequency increases with an increase in the conduction angle or the amplitude of applied voltage. Thus as the mean applied voltage is increased, causing an increase in the mean conduction angle, the circuit becomes faster. When the frequency of variation approaches that of the carrier frequency the admittance becomes zero. It is important to note that the relations given are only valid for small variations.

Transient Response

The bistable behavior of the basic ferroresonant circuit makes it suitable for a variety of practical applications.[8] In most of these applications the circuit undergoes a transient state from one stable mode to the next. Thus a knowledge of the transient behavior of the circuit is of practical importance. The dynamic equations derived in this paper are believed to contain enough information necessary for the determination of the transient response of the circuit. The author wishes only to outline a procedure for the determination of the transient response in the hope that it may be considered by other investigators for thorough treatment.

The dynamic behavior of the idealized series ferroresonance circuit is described by equation 15. For transients due to step changes in applied voltage, $\Delta E_n{}^*=0$, while $(E_n{}^*-E_{DC})$, is equal to the step change in the value of the applied voltage. Denoting this change in the applied voltage by $E^{*\prime}$, equation 15 may be written as

$$K_3\Delta RI_n+[2K_3-\pi\tan\phi]\times\left[RI_n-\frac{[2E_{DC}-4\times10^{-8}\times AB_cN\omega K_2]}{[2K_3-\pi\tan\phi]}\right]=2E^{*\prime} \quad (18)$$

The quantity in the large bracket represents the change in current from its initial value, that is, the half-cyclic average value of the current before the step change in the applied voltage is made. Denoting the change in current by $I_n{}'$, equation 18 may be written as

$$K_3\Delta RI_n{}'+[2K_3-\pi\tan\phi]I_n{}'=2E^{*\prime} \quad (19)$$

The solution of this equation is the

sum of a steady-state and a transient term which is of the form:

$$I_n' = I_{ss}' + D\left[1 - \frac{[2K_3 - \pi \tan\phi]}{K_3}\right]^n \qquad (20)$$

where

I_n' = the half-cyclic change in the average value of current referred to the initial value before the step change is made
I_{ss}' = the change in the average value of the current referred to the initial value when the transient ceases and steady-state operation is reached
K_3 = defined in equation 9
D = a constant depending on the initial conditions

By substituting the condition $I_n'=0$, at $n=0$, in equation 20, the constant D may be evaluated, and thus the equation for current becomes

$$I_n' = I_{ss}\left[1 - \left(1 - \frac{2K_3 - \pi \tan\phi}{K_3}\right)^n\right] \qquad (21)$$

From this equation, the transient behavior of the circuit may be determined. As an example, let the applied voltage be increased suddenly from its initial value E_1^* to E_2^*. Assuming operation in the stable region, the corresponding change in the half-cyclic average value of current flowing in the circuit undergoes a transient reaching a final steady-state value.

The transient may be calculated as follows:

1. Find the value of I_1 and I_2 corresponding to E_1^* and E_2^* from the steady-state equations 5 and 6, expressed graphically in Fig. 7(B). The difference (I_1-I_2) is equal to I_{ss}.
2. Find the value of α_c corresponding to E_1 from equation 5, expressed graphically in Fig. 7(A). Substituting the value of $\alpha_c=\lambda$ in equation 9, the corresponding value of K_3 may be calculated.
3. Knowing I_{ss} and K_3, the change in current during each half-cycle may be obtained by substituting consecutive integer values of n in equation 21, starting from $n=1$.

The value of K_3, determined in step 2, varies continuously with conduction angle. However, for small changes, K_3 may be taken as a constant without appreciable error. For large changes, this assumption may cause an appreciable error. To avoid this, the method of calculation must be so modified that the transient is divided into several stages according to the degree of accuracy required. The final value of one stage is taken as the initial value of the next stage.

A similar procedure may be used to determine the transition time between stable state, if equation 6 is assumed to be valid. As an example, let the voltage be increased from zero to the critical firing voltage at which the current rises to a fixed value determined by the phase angle as seen from Fig. 7(B). Just before the critical firing voltage is reached, the current is zero. The transient associated with the variation of current from zero to its final value I_{ss} may be determined by assuming this change to take place in several stages as before. The value of α_c corresponding to each stage may be determined from equation 6. By substituting $\alpha_c=\lambda$ in equation 9, the corresponding value of K_3 may be determined. It should be noted that the quantity

$$[1-(2K_3-\pi\tan\phi)/K_3]$$

in equation 21 exceeds unity since the circuit is unstable during the transition. This is different from the previous case for operation in the stable region.

Theory of Ferroresonance

The circuit analyzed is the simplest circuit exhibiting unstable and multivalued behavior. The behavior of the circuit is important from both theoretical and practical considerations and has interested engineers for a long time. In a paper by Suits,[6] and subsequent papers analyzing the same circuit, for the less ideal case of gradually varying inductance, the jump phenomenon occurring in this circuit is attributed to an instantaneous resonance taking place when the nonlinear inductance (variable with the amplitude of the forcing function) reaches such a value as to resonate with the fixing capacitance. This is probably the reason for using the term ferroresonance to refer to the jump phenomenon. This phenomenon still exists, and becomes even more prominent, when the saturation characteristic of the nonlinear reactor becomes sharper as indicated in Fig. 3, where curve *a* corresponds to an ideal magnetization characteristic consisting of regions of zero and infinite inductance as indicated in Fig. 2. Since such a reactor having an instantaneous inductance of either zero or infinity could not resonate at any instant with a finite capacitance, such an explanation of the phenomenon is thus contradicted. Another explanation of the phenomenon is attributed to an average resonance between the nonlinear reactor and the fixed capacitance. This view may be contradicted from an examination of equations 11 and 12, describing the dynamic behavior of the circuit on an average basis. Since these equations are only of the first order, there is no possibility of an average resonance.

An accurate explanation of the ferroresonance phenomena may be based on the following two concepts.

NEGATIVE RESISTANCE

The concept of the negative resistance to explain the unstable behavior of a series ferroresonance circuit has been used in the literature. However, no direct mathematical proof of the existence of a negative resistance has yet been given, except indirectly through resort to graphical plots of equations in which the presence of negative slopes are detected. In this paper this concept is varified mathematically from an investigation of the stability of equations 11 and 12 directly, without the need for graphical plots.

FEEDBACK

The concept of feedback to explain unstable behavior was pointed out by the author in reference 4, where an ideal series ferroresonant circuit was analyzed for the case of sinusoidal applied voltage and the instability was attributed to the charge on the capacitor acting as a voltage source helping the applied voltage to saturate the core. Thus, at the instant when the critical firing voltage is just exceeded, the reactor fires and the capacitor charges. The constant voltage V across the capacitor during presaturation intervals in helping the applied voltage will reduce the presaturation intervals; consequently, the conduction intervals are increased. The increase in conduction angle causes more current to flow which, in turn, raises the presaturation voltage across the capacitor. Such an accumulative action, which may be considered as voltage feedback, was used to explain the reason for instability. However, due to the complexity of equations involved, it was not possible to derive the corresponding feedback loop. The simplicity of equations deduced in this paper on the basis of a rectangular excitation makes such a representation possible. This is indicated by the nonlinear closed-loop diagram of Fig. 10, which corresponds to equations 4 and 5. The nonlinear transfer functions depend on the conduction angle which is considered as the output, while the nondimensional applied voltage $E/AB_cN\omega$, is considered as the input. The feedback quantity is $V/AB_cN\omega$, which is the nondimensional value of the voltage across the capacitor during the nonconduction intervals. At $\phi=0$, the gain of feedback loop

$$[1-\epsilon^{-\tan\phi\alpha_c}]/10^8(\pi-\alpha_c)\alpha_c$$

is zero; that is, there is no feedback. For this case the ratio of input to output is given by

$$\frac{\alpha_c}{E^*/AB_cN\omega}=\frac{10^8}{2}\,[\pi-\alpha_c]\alpha_c$$

This equation may be written as

$$\frac{E^*}{AB_cN\omega}=\frac{2}{10^8}\,\frac{1}{[\pi-\alpha_c]}$$

This is the relationship between the applied voltage and conduction angle for a reactor with pure resistive load and rectangular forcing function. This may be proved by letting $\phi=0$ in equation 5. The addition of capacitance is, in effect, equivalent to the closing of the feedback loop. Thus, as ϕ is increased, the gain of the feedback loop increases. Consequently, $V/AB_cN\omega$, which is the feedback voltage, increases. This voltage which appears across the capacitance during the nonconduction intervals is added to the applied voltage; that is, the feedback is positive.

Appendix. Nomenclature

A = effective core area (measures area × stacking factor of the core material), square centimeters
B_c = critical value of flux density at which the core saturates, gauss
C = load capacitance, farads
E^* = amplitude of a constant rectangular alternating voltage applied to the ferroresonant circuit, volts
E_{DC}^* = mean amplitude or initial amplitude of a rectangular alternating voltage of varying amplitude, volts
$E^{*\prime}$ = variation of the amplitude of a rectangular alternating voltage from the mean average amplitude E_{DC}^* or from a given initial amplitude, volts
E_n^*, E_{n+1}^*, E_{n+2}^* = variation of the amplitude of a rectangular alternating voltage from the mean average amplitude E_{DC}^* or from a given initial amplitude, during n, $n+1$, $n+2$ half-cycle, volts
e = instantaneous value of voltage applied to the ferroresonant circuit, volts
f = frequency of the voltage applied to the ferroresonant circuit, cycles per second
I = half-cyclic average value of current flowing in the ferroresonant circuit, amperes
I_n, I_{n+1}, I_{n+2} = half-cyclic average value of current flowing in the ferroresonant circuit during n, $n+1$, and $n+2$ half-cycle of the applied voltage, amperes
I' = variation of the half-cyclic average value of the current flowing in the ferroresonant circuit from the mean average value or from a given initial value, amperes
I_n', I_{n+1}', I_{n+2}' = variation of the half-cyclic average value of current flowing in the ferroresonant circuit from the mean average value or from a given initial value, during n, $n+1$, $n+2$ half-cycle of the applied voltage, amperes
i = instantaneous value of current flowing in the ferroresonant circuit, amperes
i_n, i_{n+1}, i_{n+2} = instantaneous value of the current flowing in the ferroresonant circuit during n, $n+1$, $n+2$ half-cycle of the applied voltage, amperes
N = number of turns of the reactor of the ferroresonant circuit
Q = constant capacitor charge during nonconduction intervals, coulombs
Q_n, Q_{n+1}, Q_{n+2} = constant charge across the capacitor during the nonconduction intervals of n, $n+1$, $n+2$ half-cycle of the applied voltage, coulombs
q = instantaneous value of capacitor charge, coulombs
q_n, q_{n+1}, q_{n+2} = instantaneous value of capacitor charge during n, $n+1$, $n+2$ half-cycle of the applied voltage, coulombs
R_w = resistance of the reactor winding of the ferroresonant circuit, ohms
R_L = load resistance of the ferroresonant circuit, ohms
R = effective resistance of the basic ferroresonant circuit, $R=R_w+R_L$, ohms
V = constant capacitor voltage during nonconduction intervals
V_n, V_{n+1}, V_{n+2} = constant voltage across the capacitor during the nonconduction intervals of n, $n+1$, $n+2$ half-cycle of the applied voltage
v = instantaneous capacitor voltage
α_c = conduction angle for half a cycle of the alternating voltage applied to the ferroresonant circuit during which conduction is maintained at steady-state operation (defined in Fig. 5), degrees or radians
$\alpha_{c(n)}$, $\alpha_{c(n+1)}$, $\alpha_{c(n+2)}$ = conduction angle for the n, $n+1$, $n+2$ half-cycle of an a-c rectangular voltage applied to the ferroresonant circuit during which conduction is maintained for transient operation (defined in Fig. 5), degrees or radians
β = instantaneous value of flux density, gauss
λ = steady-state operation or mean conduction angle α_c of the ferroresonant circuit about which transient variation is assumed to take place, degrees or radians
τ = time of duration of half a cycle of the applied voltage, seconds
ϕ = phase angle of the basic ferroresonant circuit, $\phi=\tan^{-1}(1/\omega CR)$, degrees or radians
Φ = total flux
ω = angular frequency of the rectangular applied voltage to the basic ferroresonant circuit, radians per second

References

1. CRITICAL CONDITIONS IN FERRORESONANCE, P. H. Odessy, E. Weber. *AIEE Transactions*, vol. 57, Aug. 1938, pp. 444–52.

2. SIMILITUDE OF CRITICAL CONDITIONS IN FERRORESONANT CIRCUITS, W. T. Thomson. *Ibid.*, vol. 58, Mar. 1939, pp. 127–30.

3. RESONANT NONLINEAR CONTROL CIRCUITS, W. T. Thomson. *Ibid.*, vol. 57, Aug. 1938, pp. 469–76.

4. ANALYSIS OF INSTABILITY AND RESPONSE OF REACTORS WITH RECTANGULAR HYSTERESIS LOOP CORE MATERIAL IN SERIES WITH CAPACITORS, J. T. Salihi. *Ibid.*, pt. I (*Communication and Electronics*), vol. 75, July 1956, pp. 296–307.

5. SINUSOIDAL AND RELAXATION OSCILLATIONS SUSTAINED BY NONLINEAR REACTANCES, R. S. Mackay. *Journal of Applied Physics*, New York, N. Y., vol. 24, no. 9, Sept. 1953, pp. 1163–67.

6. CAPACITIVE LOADING OF NONLINEAR MAGNETIC CIRCUITS, J. T. Salihi. *Ph.D. Thesis*, University of California, Berkeley, Calif., 1958.

7. STUDIES IN NONLINEAR CIRCUITS, C. G. Suits. *AIEE Transactions*, vol. 50, June 1931, pp. 724–36.

8. MATHEMATICAL ANALYSIS OF NONLINEAR CIRCUITS, A. Boyajian. *General Electric Review*, Schenectady, N. Y., vol. 34, Sept. 1931, pp. 531–37.

9. TRANSIENTS IN LINEAR SYSTEMS (book), Murray F. Gardner, John L. Barnes. John Wiley & Sons, Inc., New York, N. Y., 1948.

10. INTERESTING NONLINEAR EFFECTS, R. S. Mackay. *Journal of Applied Physics*, vol. 24, no. 3, Mar. 1953, pp. 311–13.

◆

Discussion

R. H. Dennard (International Business Machines Corporation, Yorktown Heights, N. Y.): As Mr. Salihi points out, the conditions he assumes for his analysis should be considered as a limiting case of a general problem. His use of the word "idealized" should not be interpreted to imply any inaccuracy in the analysis which is perfectly valid under certain conditions. Perhaps those conditions should be outlined more clearly. The assumption of zero inductance for the saturated reactor is valid when the product R^2C is sufficiently large as compared with the actual differential inductance for the saturated condition. Under this condition the second-order differential equation for the interval of saturation gives effectively the same solution as the assumed first-order equation. The assumption of zero magnetizing current for the unsaturated reactor is valid when the capacitance is large enough so that it does not discharge appreciably due to the flow of magnetizing current. In addition, the circuit resistance must, of course, be small enough to have negligible voltage drop during the unsaturated interval. An analysis for the case in which the saturated inductance and the magnetizing current are significant factors has been given in the recent literature.[1–3] The mathematical relationships become somewhat more complicated, particularly for the case of sinusoidal applied voltage, but a physical understanding of the circuit behavior is still possible.

Mr. Salihi shows for the case under consideration that points of instability deduced graphically from minimum points on steady-state curves coincide with the points obtained from a mathematical investigation of the dynamic equations, claiming that the former verifies the latter. From this, one might tend to conclude that stability could always be determined by examining the slope of plots of average current versus applied voltage. That this is not valid is shown by the following example.

Evans[1] notes a case in which circuit operation with symmetrical waveforms is superseded by an unsymmetrical operation as the applied voltage is increased. Subsequent literature[2] shows that there is no negative slope in the current-voltage relationship at the point of transition, but an investigation using difference equations verifies the experimental evidence that the symmetrical type of operation becomes unstable at that point.

REFERENCES

1. SERIES CIRCUIT INCORPORATING A NON-LINEAR REACTOR, R. P. Evans. *AIEE CP 56-683* (available on request).

2. BEHAVIOR OF THE FERRORESONANT SERIES CIRCUIT CONTAINING A SQUARE-LOOP REACTOR, R. H. Dennard. *Doctoral Dissertation*, Carnegie Institute of Technology, Pittsburgh, Pa., 1958.

3. BEHAVIOR OF THE FERRORESONANT SERIES CIRCUIT CONTAINING A SQUARE-LOOP REACTOR, R. H. Dennard. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, 1958 (Jan. 1959 section), pp. 903–11.

Gordon R. Slemon (University of Toronto, Toronto, Ont., Canada): I would like to commend Mr. Salihi for pointing out the error in applying the concept of resonance to this type of behavior. Resonance is normally associated with the cyclical transfer of energy between two energy-storage elements. In an ideal switching core there is no energy storage. Thus "ferroresonance consists of a cyclical energy transfer between a source and a capacitive element."

I would question the term "negative resistance" which the author has used. This term also arises from a doubtful analogy with the behavior of linear circuits. While there is a negative damping term in the expression for average voltage (equation 12), the fact that the average voltage is not in equilibrium does not imply the property of negative resistance.

There are other stable modes of operation of this simple circuit normally known as the subharmonic modes. I would appreciate Mr. Salihi's comments on the possibility of applying his difference equation methods to determine the regions of stability of these modes.

H. C. Bourne for Jalal T. Salihi: Dr. Salihi is now in Iraq, address unknown, and unable to write a closure to these discussions. However, I am sure that he appreciates the accurate and pertinent comments by Drs. Dennard and Slemon.

A Small High-Speed Transistor and Ferrite-Core Memory System

W. L. SHAFER, JR.
ASSOCIATE MEMBER AIEE

W. N. TOY
NONMEMBER AIEE

H. F. PRIEBE, JR.
NONMEMBER AIEE

Synopsis: Memory for computer-type systems has received extensive study and design effort in the past few years. This effort has been directed mainly at producing faster, larger, and less expensive systems in the thousand- to mega-bit sizes.

This paper describes the design and operation of a 400-bit memory for use in a time-division electronic switching system.

THE design objective of our exploratory development project was to construct a 400-bit memory system to be used in a time-division switching system where the number of circuits controlled required 20 bits to be read out of the memory simultaneously. Our principal objective was to obtain as short a read-write time as possible under the following operating conditions:

1. Temperature. Since this memory was to be a part of a small system, used in a variety of locations, temperature control was not practical. Operation over an equipment temperature range of 30 to 140 F (degrees Fahrenheit) was required.

2. Simplicity and, consequently, cost. In a small memory system the access circuitry accounts for the largest part of the cost. Therefore, simplicity in these circuits was desirable.

3. Solid-state devices. For system size and reliability it was decided that only solid-state components: transistors, diodes, and ferrite memory elements, would be used.

Choice of Memory Element

Besides the ferrite core, there are a number of memory devices, such as the twistor, which are in various stages of research or development which offer promise of high speeds. However, for this system it was decided to use the ferrite core because of its present wide acceptance as a reliable device. Of the many commercially available cores with like characteristics and switching times in the 2-μsec (microsecond) range the General Ceramics Corporation *S4-F394* core was chosen.

General Systems Block Diagram

A block diagram of the over-all system is shown in Fig. 1. The information in each row of the core array is read and written back in time sequence. The read pulse from the horizontal drive circuit reads a row of cores. The signals from the 20 cores are examined by readout detectors which discriminate between a *1* and a *0*. Immediately following, a write pulse is delivered to the same row. Prior to the write pulse the system has received information as to whether the same or dif-

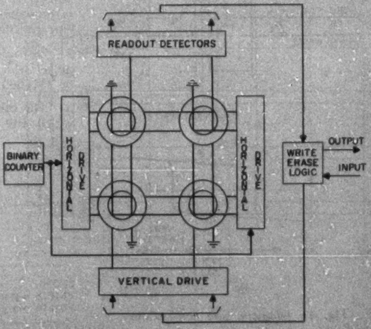


Fig. 1. Block diagram of memory systems

Paper 59-826, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted December 31, 1958; made available for printing April 22, 1959.

W. L. SHAFER, JR., W. N. TOY, and H. F. PRIEBE, JR., are with Bell Telephone Laboratories, Inc., Murray Hill, N. J.

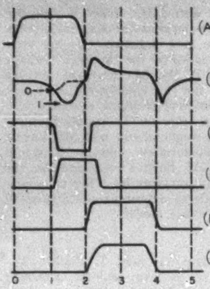
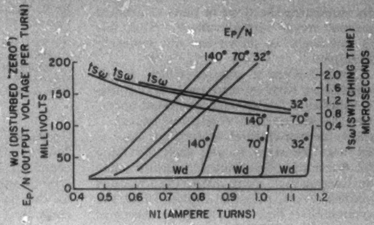


Fig. 2 (left). Memory timing

A—Horizontal read
B—Memory output
C—Memory strobe
D—Detector output
E—Vertical write
F—Horizontal write

Fig. 3 (right). Memory core characteristics, General Ceramics Corporation S4-F394

Pulse rise time, 0.3 μsec
Pulse width, 2.0 μsec



ferent information is to be written back into the memory. The write-erase logic then causes the appropriate vertical-drive circuit to produce write pulses. The coincidence of the horizontal and vertical pulses causes the selected cores to switch, thus writing the information. During the next time slot the adjacent row of cores, i.e., the next word, is subjected to the same operation. This process of reading and rewriting each word (or writing a new word) continues in a sequential fashion until all 20 words have been processed. The cycle is then repeated. The sequential operation of the memory is controlled by the binary counter which delivers 20 binary words to the horizontal drive and then recycles. The horizontal-drive circuitry acts as a stepping switch, sequentially selecting one out of the 20 rows in the memory.

The read-write operation for each word requires 5 μsec as shown in Fig. 2. This 5-μsec time slot is divided into the read and write phase by means of timing pulses from an external clock. The cycle time for processing all 20 words is 100 μsec.

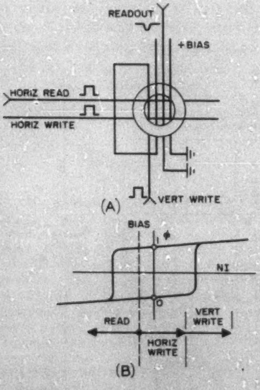


Fig. 4. Memory core connections

Memory Core

The memory-core characteristics place requirements on the access circuitry as well as determine, to a large degree, the systems operating margins. Twenty-five *S4* cores were measured and the resulting data are shown in Fig. 3. These data represent the limiting characteristics. For example, the switching-time curves show the slowest core-switching time for each of the conditions of temperature and drive. The output voltage represents the lowest output voltage for a *1* from any of the 25 cores. The walkdown (W_d) curve represents the highest read voltage from any core after a series of half-write pulses.

The W_d curves as plotted were obtained by using half the pulse current of that reported on the abscissa. Thus W_d at say 800 ma (milliamperes), is not with a disturbing current of 800 ma but of 400 ma. This is similar to actual system operation.

From Fig. 3 the unbiased operating limits can be determined. The drive-current limits are determined by three points:

1. The maximum amplitude of a series of half-write pulses at 140 F which will not cause switching of an unselected core. This current is found to be 400 ma.
2. The minimum amplitude of a full-write pulse, which will just switch the core at 30 F, is 660 ma.
3. The minimum amplitude of read pulse which will provide a satisfactory output signal, as determined from the curves, is 780 ma.

With no bias the operating points are then: read, minimum 780 ma; and one-half write, minimum 330 ma, maximum 400 ma.

Biased operation of the memory offers several advantages. When coincidence current operation is used only for writing, bias in the read direction offers the following advantages: greater variation in the half-write current, and reduction in read drive required.

Setting the bias at 300 ma the pulse current amplitudes are as follows: minimum amplitude of read, 480 ma; and one-half write, 480 ma to 700 ma.

Memory-Array Wiring

D-c bias and current pulses are applied to the memory core as determined by the wiring for the array which is shown in Fig. 4(A). Two horizontal wires are provided; one for the read-drive pulse and the other for the horizontal half write. Four vertical wires are provided; one for bias, one for the readout signal, and two for a double-turn vertical half-write pulse. A hysteresis loop of a core is shown in Fig. 4(B). The bias is in the read direction such that nominal-read, horizontal-write, and vertical-write drives are approximately equal amplitude.

Transistors

Two basic types of transistors are used throughout the memory system. One is a low-power n-p-n diffused silicon unit and the other a medium-power germanium-

Table I. Transistor Characteristics

	β at I_e, ma					$f_{\alpha b}$, mc	Vc	Td, C	Tl, C
	25	100	150	200	600				
2N560	50	55	52	16	x	50	30	0.25 mω	175
F51618 n-p-n F51619 p-n-p	50	40	38	35	20	4	30	.9 mω	80

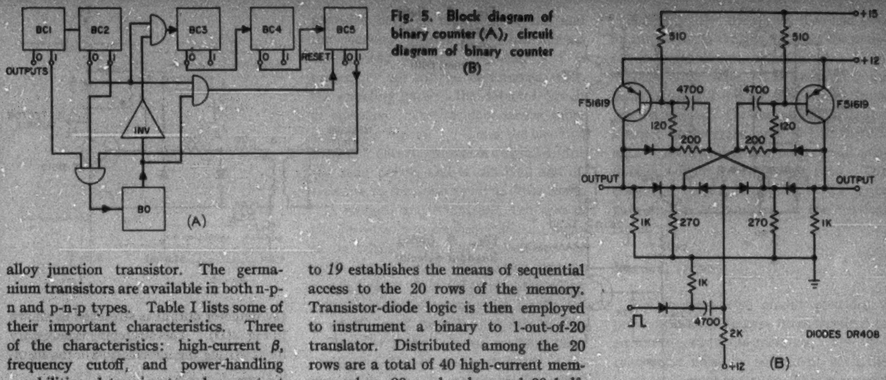

BC1
BC2
BC3
BC4
BC5
OUTPUTS
0
1
RESET
INV
BO
(A)
+15
510
510
+12
F51619
F51619
4700
4700
120
120
200
200
OUTPUT
OUTPUT
1K
270
270
1K
1K
4700
2K
+12
DIODES DR408
(B)

Fig. 5. Block diagram of binary counter (A); circuit diagram of binary counter (B)

alloy junction transistor. The germanium transistors are available in both n-p-n and p-n-p types. Table I lists some of their important characteristics. Three of the characteristics: high-current β, frequency cutoff, and power-handling capabilities, determine to a large extent the over-all systems design and performance. The number of stages necessary for a desired drive current depends on β, of course. The upper limit on the speed at which the access circuitry can be expected to perform depends on frequency cutoff and power-handling capabilities. These are not much out of line with the core performance for the design speed.

The silicon transistor is used wherever circuit design permits, and eventually it is expected that silicon will be used throughout. The vertical circuits make use of both types of transistors while the horizontal circuits are made up entirely of the germanium units.

Horizontal Drive and Binary Counter

Two basic building blocks are used to make up the horizontal-drive circuits. A 5-stage binary counter cycling from *0* to *19* establishes the means of sequential access to the 20 rows of the memory. Transistor-diode logic is then employed to instrument a binary to 1-out-of-20 translator. Distributed among the 20 rows are a total of 40 high-current memory pulses, 20 read pulses and 20 half-write pulses. The 5-stage binary counter's normal cycle of 32 is shortened to 20 by a reset circuit, or, more exactly, an advance circuit. A block diagram of the counter is shown in Fig. 5(A). The binary-counter stages are of special design to provide a good output waveform. A circuit diagram is shown in Fig. 5(B). Diodes connected between the transistor's collector and the cross-coupling capacitor provide isolation which permits an output waveform determined largely by the transistor characteristics and not by other circuit elements. Also, diodes have been connected between base and collector terminals of the transistors to limit the degree to which high-β units will go into saturation. Rise and fall times of the counter stages average about 0.1 μsec and 0.7 μsec, respectively. The longest advance time, when all stages are switching, is approximately 1 μsec. Since the transistors employed in the counter are of the medium-power variety, they are capable of 200-ma output-pulse current which minimizes the number of subsequent amplifiers.

The horizontal-drive circuits are shown in block form in Fig. 6. The first two binary counter stages operate in conjunction with a binary to 1-out-of-4 translator to operate transistor switches on one side of the memory. On the other side of the memory the five transistor switches are operated from the remaining three binary counter stages and a 1-out-of-5 translator. Thus, the 1-out-of-20 selection is accomplished in matrix form with two translations: 1-out-of-4 and 1-out-of-5; these are combined as shown in Fig. 7. Relative timing is also shown in the figure. Two translations of 1-out-of-4 are used: one is combined with the clock-read pulses and the other is combined with the write pulses.

A rather high component efficiency is achieved by this circuit arrangement. A total of 37 transistors, including those in

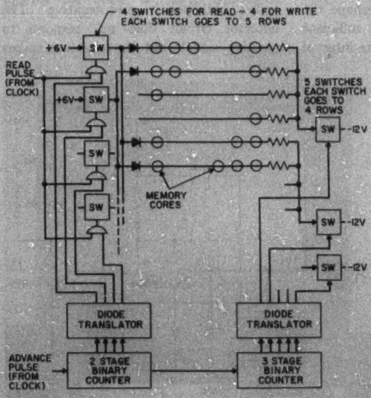

4 SWITCHES FOR READ – 4 FOR WRITE
EACH SWITCH GOES TO 5 ROWS
+6V
SW
+6V
SW
SW
SW
READ PULSE (FROM CLOCK)
5 SWITCHES EACH SWITCH GOES TO 4 ROWS
SW
-12V
SW
-12V
SW
-12V
MEMORY CORES
DIODE TRANSLATOR
DIODE TRANSLATOR
ADVANCE PULSE (FROM CLOCK)
2 STAGE BINARY COUNTER
3 STAGE BINARY COUNTER

Fig. 6 (left). Horizontal drive

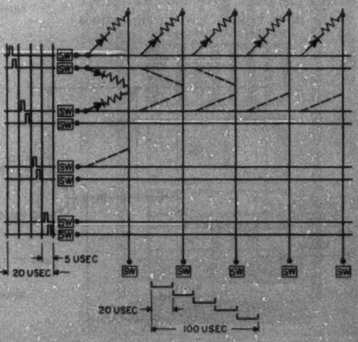

SW
20 USEC
5 USEC
20 USEC
100 USEC

Fig. 7 (right). Horizontal read-write timing

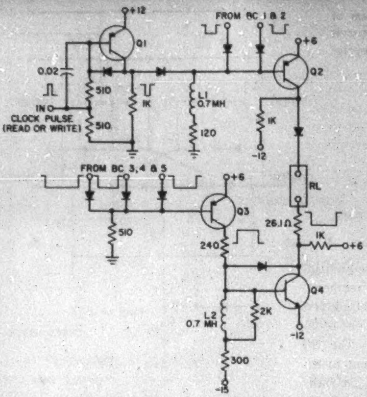

+12
Q1
FROM BC 1 & 2
+6
Q2
0.02
510
1K
L1
0.7MH
IN
CLOCK PULSE
(READ OR WRITE)
510
120
1K
-12
FROM BC 3,4 & 5
+6
RL
Q3
510
240
26.1Ω
1K
+6
Q4
L2
0.7 MH
2K
-12
300
-15

Fig. 8 (left). Typical horizontal pulse-current path

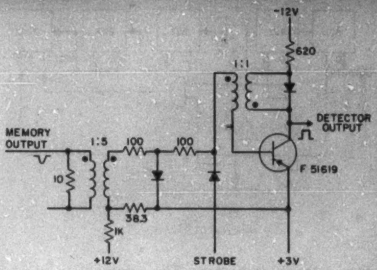

-12V
620
1:1
DETECTOR OUTPUT
MEMORY OUTPUT
1:5
100
100
10
F 51619
383
1K
+12V
STROBE
+3V

Fig. 9 (right). Readout detector

the binary counter, provide the 40 horizontal read and write pulses.

A typical pulse-current path is shown in Fig. 8. Transistor *Q1* amplifies the clock pulse and drives four gate circuits similar to that connected to the input of *Q2*. The horizontal wire through the memory is shown in *RL*. With a positive-going pulse applied to one side of the memory wire an appropriate negative-going pulse is applied to the other by a circuit such as *Q3* and *Q4*. Transistors *Q1*, *Q2*, and *Q3* are p-n-p types and *Q4* is an n-p-n type. The binary counter and *X* pulse circuits operate between ground and the positive battery while the *Y* pulse circuits operate between ground and the negative battery. Stage *Q3* provides the inversion and voltage shift necessary to connect the diode logic from the binary stages to the n-p-n transistor *Q4*.

The power dissipation and, consequently, heating, among transistors is quite evenly distributed. The duty cycle for stages such as *Q2* is approximately one-eighth and for *Q4* it is approximately one-sixth. The collector voltage of the units is kept low during periods of high currents which keeps the collector dissipation to a minimum.

Operation of the memory depends on constant-current pulses, particularly in the write interval when the coincidence of horizontal and vertical pulses is required. The required current regulation is obtained from a voltage source and suitable dropping resistor. The variation in potential drops across transistors and diodes accounts for a large percentage of the existing current variation.

Readout Detector

During the read interval, the switched cores produce an output voltage and the nonswitched cores, a shuttle or noise voltage. These waveforms are shown in Fig. 2(B). The voltage amplitude of a *0* is made up of crosstalk as well as the core shuttle voltage. The readout detector must distinguish the wanted signal from the noise and it must reshape the wanted signal. To provide sufficient delay from the read to the write interval the detector output is stretched as shown in Fig. 2(D). This stretching is necessary for rewriting the information in the memory.

The readout detector diagram is shown in Fig. 9. It is a blocking oscillator circuit with collector-to-base feedback. Since the minimum output from the memory for a *1* is 100 mv (millivolts) and the maximum noise for a *0* can be as high as 50 mv during the read interval, the detector is designed to have a threshold of 60 to 90 mv. The input transformer provides a better impedance match between the memory and the blocking oscillator. The transistor is normally biased in the cutoff condition, and when the signal from the memory exceeds its threshold, an output is produced. The output-pulse width depends largely on the value of the inductance of the feedback transformer, the β of the transistor, and the circuit resistances.

In addition to the noise produced by the shuttling of the cores during the read interval, there is also a large noise caused by the coupling from the vertical-write wires. The noise from the vertical-write pulse is in the form of spikes, both positive and negative. The detector circuit does not respond to positive signals but the negative spikes

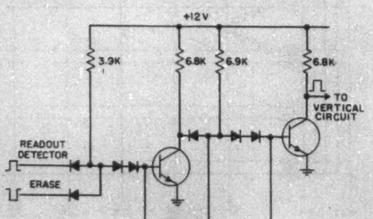

+12V
3.9K
6.8K
6.9K
6.8K
TO VERTICAL CIRCUIT
READOUT DETECTOR
ERASE
WRITE
47K
47K
-12V

Fig. 10 (left). Write and erase gates

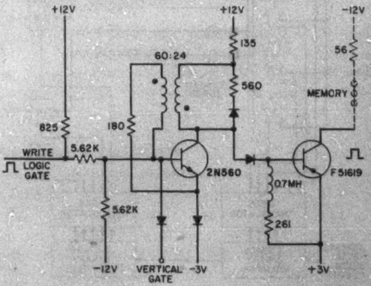
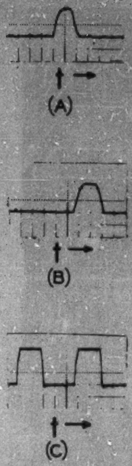

+12V
+12V
-12V
135
56
60:24
560
MEMORY
825
180
WRITE
LOGIC GATE
5.62K
2N560
F51619
0.7MH
5.62K
261
-12V
VERTICAL GATE
-3V
+3V

Fig. 11 (right). Vertical half-write-drive circuit



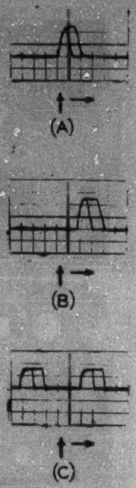


Fig. 12 (left). Nominal read and write current pulses

A—Horizontal read, 250 ma/cm, 1 μsec/cm
B—Horizontal write, 250 ma/cm, 1 μsec/cm
C—Vertical write, 90 ma/cm, 1 μsec/cm

Fig. 13 (right). Nominal and limiting read and write current pulses

A—Horizontal read, 250 ma/cm, 1 μsec/cm
B—Horizontal write, 250 ma/cm, 1 μsec/cm
C—Vertical write, 180 ma/cm, 1 μsec/cm

are in the same direction as the desired signal and about twice as large in magnitude. The negative spike occurs at the trailing edge of the write pulse just ahead of the read output. To guard against the noise, a 1-μsec interval is provided between read-write cycles and the memory output is strobed during the readout interval. Whether a *1* is present or not will be determined when the enabling gate in the detector circuit is operated. The memory-strobe pulse is shown in Fig. 2(C). Thus, the blocking oscillator detects the wanted signal during the read period and is disabled and insensitive to the noise signal at other times.

A second, and important, function of the detector is to produce sufficient delay in its output such that it overlaps the write interval and information can be recirculated. This delay is obtained by extending the detector-pulse width beyond the read period as shown in Fig. 2(D). To make sure of readout pulse stretching, the blocking oscillator is designed to have its natural-pulse width greater than the clock-pulse width and the clock pulse is then used to turn the blocking oscillator off.

Write and Erase Logic Gates

Writing new information and erasing the old is carried out by the write and erase logic gates. The circuit is shown in Fig. 10. The gates are the connecting links between the readout detector and the vertical-drive circuit. When a write pulse appears at the input of the write logic gate, an output is produced by the vertical-drive circuit. The coincidence of the vertical drive and the horizontal drive writes the information into the memory. When the information is read, it is also written back immediately if there is no erase pulse present. The write logic gate has two inputs, one is the output of the erase gate and the other is an external write pulse; the presence of either input will produce an output from the write gate.

Similarly, the erase gate also has two inputs. One is the output from the readout detector circuit, and the other is an external-erase pulse. The presence of an erase pulse will inhibit any output from the erase gate. On the other hand, the absence of the erase pulse in conjunction with a readout-detector output will produce an output from the erase gate. The erase-gate output is applied to the write gate and is transmitted to the vertical circuit.

Vertical Drive

The vertical-drive circuit provides a pulse of half the ampere turns required to switch a core in the memory. The current pulse must be coincident with the horizontal-write current. This is assured by letting a clock pulse determine the timing. The circuit is shown in Fig. 11. The first stage is similar to the readout-detector circuit in that it uses a blocking oscillator. The output from the logic gate begins during the read interval and lasts into the write interval. However, the blocking oscillator is kept cut off by the negative potential applied during the absence of the vertical-gating pulse. During the vertical-gating pulse, which is positive, the gating diode is cut off. An output from the logic gate is then allowed to trigger the blocking oscillator. The transistor will stay in the ON condition

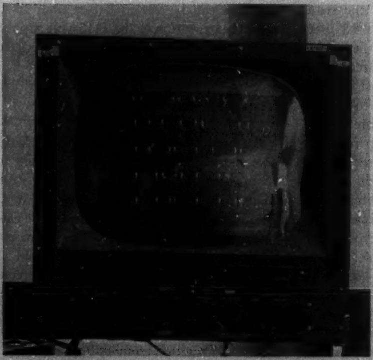


Fig. 14 (left). Visual display of memory contents

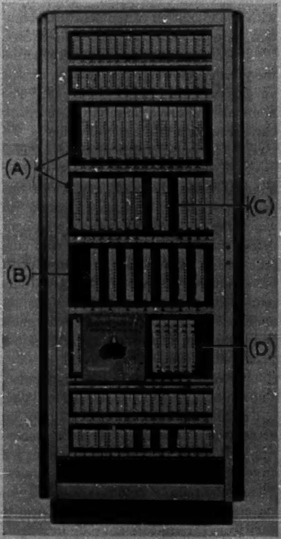


Fig. 15 (right). Memory system

A—Detectors and vertical drive
B—Horizontal read-write
C—Memory
D—Clock

until the gating pulse returns to the negative potential. Thus, in a manner similar to the readout detector, the vertical-output-pulse width in controlled by the gating pulse.

The second vertical-write stage is simply a transistor switch that provides 270 ma of current for driving the memory. The inductance in base circuit helps to turn the transistor ON and OFF rapidly and the rise and fall time of 0.2 μsec are obtained. Stretching of the pulse due to storage time is minimized by the large initial reverse currents provided by the inductor.

Over-all Operating Characteristics

1. Limiting conditions. Limiting conditions associated with reading and writing determine the range of successful memory operation. To have some measure of tolerances of the system, limiting pulse position, width, and amplitude were determined. Since the clock was designed with sufficient flexibility so that the timing and pulse width could be adjusted, easy testing of the memory was possible. The following items were investigated and the results will be discussed: (*a*). minimum horizontal read-pulse width, (*b*). minimum readout detector strobing-pulse width, (*c*). minimum coincidence between horizontal and vertical write pulse, (*d*). minimum read-write time, (*e*). maximum supply-voltage variation, and (*f*). high- and low-temperature operation. These tests, with the exception of temperature operation, were performed at room temperature.

2. Read pulse. Photographs of actual waveforms showing the timing of the various pulses are shown in Fig. 12. The nominal read-drive pulse is 600-ma peak current with a rise and fall time of 0.4 μsec. The pulse width at the base is 2 μsec. Keeping the other pulses at their nominal value, the read-pulse width was decreased. The limiting condition in which the readout detector can no longer detect the signal was when the read current looked like one-half sine wave with base width of 1.2 μsec and the peak current of 500 ma as shown in Fig. 13(A). It appeared that a 600-ma 2-μsec read pulse had sufficient margin to guard against circuit variations.

3. Readout-detector strobing pulse. The strobe pulse samples the memory output during the read interval. It can be made as small as 0.3 μsec in width occurring at the peak of the memory output for any one particular core in the memory. The strobe pulse is common to all detector circuits. Since the memory outputs vary from core to core not only in magnitude but also in waveshape, the strobe pulse must be at least 1 μsec wide in order to sample all the memory outputs.

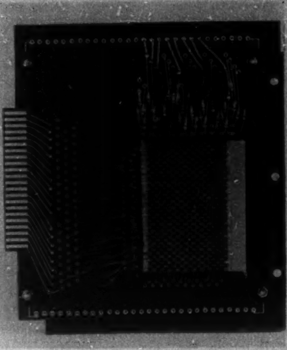


Fig. 16. Core memory array

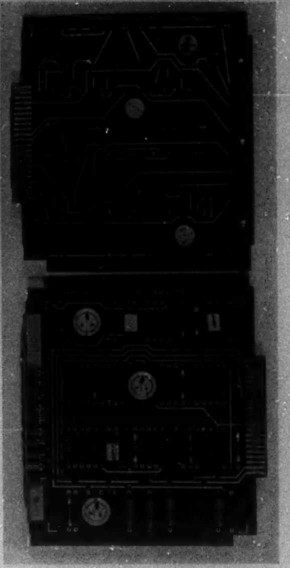


Fig. 17 (right). Vertical circuit component and wiring card

4. Coincidence of horizontal- and vertical-write currents. The results show that the coincidence between horizontal- and vertical-write-current pulses is not as critical as was expected. Their nominal pulse width, across the base, is 2.6 μsec as shown in Figs. 12(B) and (C). The vertical pulse was then set 0.7 μsec earlier and the horizontal pulse 0.7 μsec later than the nominal timing, and the memory performed satisfactorily. With coincidence of both vertical- and horizontal-write pulses the limiting-pulse width was 1.2 μsec as shown in Figs. 13(B) and (C). Beyond this, the memory can no longer recirculate its information.

5. Read-write time. The information just discussed indicates that as far as the memory element is concerned the read-write time can be decreased to 2.2 μsec. However, the outputs from the access circuitry vary from circuit to circuit such that 3 μsec is about the shortest possible read-write time that appears to be obtainable at room temperature. The present system was speeded up reducing the time slot from 5 μsec to 4 μsec to provide an indication of the margin. Time constants in the binary counters were chosen on the basis of 5-μsec time slot so that speeding up the clock beyond a 4-μsec interval was not possible. However, the operating with a 4-μsec time slot verifies the expected margin in the system.

6. Voltage tolerance. Large variation in supply voltage is a good measure of margin in the system. Seven different supply voltages are used in the drive circuits. Each voltage can be varied better than ±10%. The system also performed satisfactorily when all the supply voltages were set 10% higher and 10% lower than their nominal values, which represented their worst combinations.

7. Temperature performance. The memory system consisting of the horizontal-access circuits, the core array, and vertical-access circuits was tested at two temperatures, 30 F and 140 F. The system performed satisfactorily at these temperatures. In addition all the supply voltages can be varied better than 10% with the exception of the −12-volt supply at 30 F. It can only be varied between −11.2 volts and $>|-15 \text{ volts}|$.

Visual-Display Unit

To facilitate testing and also to check the operation of the memory system a monitoring device was constructed and placed in operation with the memory. The unit provides a visual display of the information circulating in the memory system. The bits are plotted on *X* and *Y* co-ordinates on a cathode-ray tube screen in direct correspondence with the core locations. Fig. 14 shows 40 bits distributed throughout 5 time slots as is obtained in an actual memory application.

Conclusions

The 400-bit memory system as installed in a practical application is shown in Fig. 15.

The wired-memory array as it is assembled on a printed-wire board is shown in Fig. 16.

The vertical circuits are arranged with a readout detector, write-erase logic and a vertical-write stage all on one printed board shown in Fig. 17. Twenty such vertical units are used in the system as described.

A high-speed ferrite-core memory system has been designed to provide reliable operation with voltage tolerances of ±10% over a temperature range of 30 to 140 F. A life test is presently being conducted on the system, and at the time of writing, it has been operating successfully without any component failures for more than 3,600 hours. The high degree of reliability has more than met the expected design objective.

Status of Sampled-Data Systems

E. I. JURY
MEMBER AIEE

Synopsis: This paper briefly describes in general terms procedures for analysis and methods of synthesis of sampled-data systems and indicates present areas of research and some of the important problems to be tackled in this field.

In particular, the following items are discussed:

1. Methods of analysis of linear sampled systems including the z-transform, frequency response, impulsive response, and difference equation methods.
2. Procedures for design of sampled-data systems or continuous-control systems employing digital computers. Both time and frequency-domain methods are described.
3. Types of physical implementation of discrete and continuous compensation networks.
4. Sampled-data systems with finite pulse width and pulse-modulated feedback systems and methods for their analyses.
5. Experimental study of a simulated system on a specially built computer.

This paper concludes by describing the present research activities in this field, enumerating some of the problems to be tackled, and suggesting areas for future research in this increasingly important field.

WITH THE complexities that have arisen in the area of automatic radar-tracking systems and with the advent of the use of digital computers in control systems, a considerable amount of activity has arisen in the field of sampled-data systems in recent years. In the course of investigating these systems several analytical techniques for analysis and design have been developed. The application of these techniques to fields other than sampled data is becoming increasingly important. For example, many economic systems, such as inventory control and production control, can be handled as sampled-data systems and hence the aforementioned analysis and design techniques can be applied to these systems. Furthermore, the purposeful introduction of sampling in systems to gain advantage in cost and weight of equipment is drawing considerable attention in industry.

The author wishes in this short paper to indicate some of the progress made in this field and some of the difficult problems yet to be investigated. For brevity, however, the author has found it necessary to concentrate on one aspect of linear sampled-data systems, and thus mentions only methods of analysis, methods of design, implementation of physical components, and finally, progress made in the pulse-modulated feedback systems. Many areas of research and progress have not been extensively discussed in this paper, such as statistical design, time-varying systems, nonlinear sampled-data systems, and variable sampling-rate systems which constitute much of the present activity in this field.

Applications of the techniques of sampled data to other fields, such as network and circuit theory, design of digital computers, numerical analysis, approximation problems to economic and biological systems, magnetic pulse-modulating systems, telemetry systems, information theory, and simulation of systems, constitutes a major area of progress made in this field, and should not be overlooked and, indeed, deserves much more investigation.

For further references on this subject, it is strongly recommended that the reader consult the two complete bibliographies published recently.[1,2]

Methods of Analysis

Sampled-data systems can be analyzed for stability, and their response to specific aperiodic inputs may be studied using any of the following methods: difference equation method, impulsive response method, frequency response method, or the z-transform method. These methods are discussed in detail in the literature and although each has its advantages and disadvantages, it appears from the trend of recent publications that the z-transform method is more promising.

DIFFERENCE EQUATION AND STATE VARIABLE METHODS

Sampled-data systems can be analyzed for stability and their response at sampling instants may be determined by sets of difference equations with constant coefficients.[3] For obtaining the total response, these difference equation coefficients become periodic functions. For sampled-data systems with periodically time-varying sampling rates, these equations become linear difference with periodic coefficients. Operational solutions of such equations are discussed in the literature.[4]

The state variable method[5] is based on the concept of introducing a set of numbers (called state variables) which contain all the past information of the system and from which one can evaluate the future behavior of the system for any input. By using vector notations for the state variable one can obtain the stability conditions and the output response of the sampled-data systems in terms of matrices.

IMPULSIVE RESPONSE METHOD

This method actually is a step-by-step method of analysis of sampled-data systems.[6] It has the advantage that the total response can be obtained, but suffers from computational labor and loss of insight to the behavior of the more complicated systems. It is based on the assumption that the finite pulses which excite the

---

Paper 58-1182, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26-31, 1958. Manuscript submitted July 14, 1958; made available for printing May 22, 1959.

E. I. JURY is with the University of California, Berkeley, Calif.

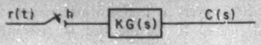


Fig. 1. Typical sampled-data system with finite pulse width

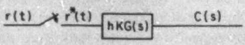


Fig. 2. A mathematical representation of system in Fig. 1

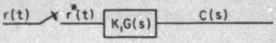


Fig. 3. Typical sampled-data system

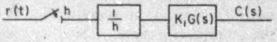


Fig. 4. Equivalent practical system of Fig. 3

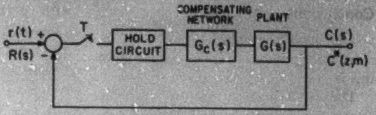


Fig. 5. Continuous network compensation of sampled-data feedback systems

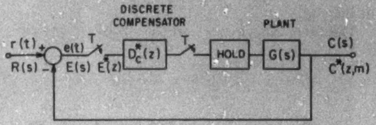


Fig. 6. Discrete compensation of sampled-data feedback systems

system are approximated by impulses or delta functions.

Frequency Response and Signal Flow Methods

This method employs the conventional s-plane techniques.[7] However, since the exact description of sampled-data systems involves mixed transfer functions in ϵ^{Ts} and s, approximation techniques are necessary for this approach. The approximation involves the reduction of the polynomial of ϵ^{Ts} into a rational function of s and by doing so, the conventional techniques developed for continuous systems can be applied to sampled-data systems.

The signal flow method is a technique based on replacing the sampler by a signal source[8] to obtain input-output relationships for sampled-data systems. This technique has been extended towards obtaining a general input-output relationship for multisampled-loop sampled-data systems.[9]

z-Transform Method

This method[10,11] with its extension, the modified z-transform,[12–15] offers an exact method for analyzing single and multirate,[16] and periodically time-varying sampling rate,[17,18] sampled-data systems. It has the advantage that the output can be expressed in a closed form Stability studies are readily effected with this approach. The z-transform method is essentially one that transforms the difference equations into algebraic equations in terms of (ϵ^{Ts}) or (z). This method has a limitation in that it is based on the approximation of finite pulses by impulses. The extension of this method into finite pulsed systems is referred to as the p-transform[19] method which will be discussed in the section on sampled-data systems with finite pulse width.

Mathematical Equivalent of Sampled-Data System

Actual sampled-data systems operate with information of finite pulse width, with these finite pulses being fed into a linear system as shown in Fig. 1. The mathematical description of the sampler using the z-transform method is based on approximating the finite pulses by impulses whose areas are equal to the input amplitude at the sampling instants. This is only valid if the gain of the original system is multiplied by h. Furthermore, this approximation is valid only if h is small compared to the dominant time constant of the system, usually between 1/5 to 1/10. The equivalent mathematical model of Fig. 1 is shown in Fig. 2.

The two systems have almost identical outputs on the condition that: 1. $KG(s)$ has a continuous impulse response, and 2. the pulse width h is small compared with the dominant time constant of $KG(s)$. Experimental results indicating the effect of varying the pulse width for one and two time-constant sampled-data systems are introduced in the experimental study section.

In most papers on sampled data, the authors have been concerned with sampled data as such, as shown in Fig. 3. However, the experimentalist must recognize that this system can be closely approximated by the physically realizable system shown in Fig. 4, in which case a division by h is necessary. Thus, the preceding discussion shows that division or multiplication by h depends on the kind of analogy desired.

In most sampled-data systems, a sort of hold-circuit is used which converts the very short pulses (impulses) into pulses usually of constant amplitude for the total sampling period. In the cases where the hold-circuit is considered "ideal" the z-transform method is used as an exact description of the physical sampled-data system.

Furthermore, if rectangular pulses with finite width are used, then it can be shown that when the pulse width becomes small the z-transform is obtained provided the gain of the system is multiplied by the pulse width, and then when the pulse width approaches the sampling period (staircase shape) the gain of the system is unchanged. This method has been successfully applied by Zypkin.[15]

Synthesis of Sampled-Data Control Systems

In general, synthesis procedures can, on the whole, be divided into two categories:

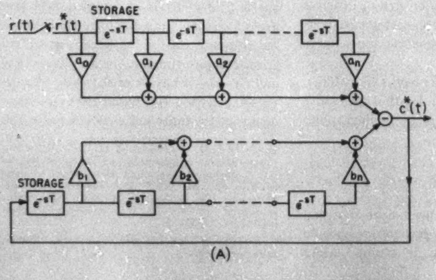
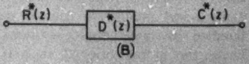


Fig. 7. Discrete compensator

A—Programing
B—Block diagram



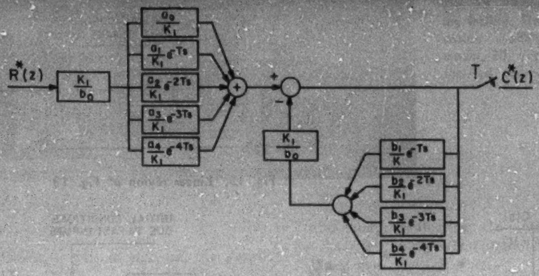


Fig. 8. Delay line of discrete compensators

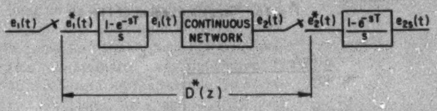


Fig. 9. Pulsed R-C network of discrete compensators

1. Synthesis with continuous compensators such as linear lumped-parameter networks.

2. Synthesis with discrete compensators, single or multirate, such as digital computers, delay lines, and pulsed networks.

As can be seen from the nature of classifications 1 and 2, synthesis procedures are somewhat different in each case.

In discussing the continuous network compensation method, it will be assumed for simplicity that series compensation, as shown in Fig. 5, is the desired form; however, minor loop compensation could be advantageously used depending on the system requirements.

The methods used for determining the transfer function $G_c(s)$ in terms of poles and zeros in the s-plane are the following:

Frequency response methods. These methods in turn could be subdivided into approximate and exact methods. The approximate methods are mainly Linvill's procedures[7] of Nyquist locus shaping in the $\Sigma KG(s)$ plane and Sklansky's z–s approximations.[30] The exact methods[18] are based on loci-shaping in the $KG^*(z)$ plane which shows the effect of $KG(s)$ poles and zeros on the sampled-data transfer function.

The frequency response methods, similar to the continuous case, are employed to yield a satisfactory value for frequency overshoots and bandwidths, which in turn influence the transient and steady-state behavior of the system.

Time-Domain methods. These methods are based on shaping the transfer function in the z-plane to yield a satisfactory response to aperiodic inputs or to certain error functions. From the over-all transfer function one can determine the z-transform of the compensating element either approximately or exactly depending on the rigid requirements imposed. Several methods[21] are proposed, mainly approximate, but they yield a realizable transfer function of the continuous compensator depending on the response requirements imposed.

The use of discrete compensators as shown in Fig. 6 simplifies the methods of synthesis considerably in view of freeing the compensator transfer function from the plant transfer function. Thus, the techniques developed for such procedures are tackled in the $z=\epsilon^{Ts}$ plane.

Similar to the continuous network synthesis both frequency and time-domain procedures are employed.

Frequency methods. These are divided into two categories: 1. shaping of the Nyquist diagram[32] in the KG^*z plane; and 2. shaping of the Bode diagram through use of the bilinear transformation:[33]

$$z=\frac{w+1}{w-1} \tag{1}$$

The frequency methods as proposed are exact and yield satisfactory results; however, frequency shaping mainly improves the response at the sampling instants. The modified z-transform can be readily employed in the frequency domain, wherein a certain measure of ripple can be estimated.

Time-domain methods. There are several methods proposed which are outlined in the following.

1. The time-domain synthesis method based on the forcing of the roots of the characteristic equation of the system to be at the origin in the z-plane[11,34] or at a particular parameter α between zero and unity, known as the staleness factor.[35]

2. The method of Bergen and Ragazzini[36] in designing the discrete compensator to yield a certain ratio of finite polynomials in Z which represents the transfer function in the z-plane.

3. Methods based on the modified z-transform to insure certain continuous responses, either dead beat or otherwise.[31,34,37]

4. Methods based on minimizing the total-square error or other error criteria.[28]

5. Root-locus method, based on relating the effect of pole and zero locations of the discrete compensator to the total response.[29,30]

This discrete compensator method, either in the frequency domain or in time-domain, can also be applied where digital computers are used in continuous systems, in which case T, the sampling period, could be changed.

It may be noted here that the use of digital elements in control systems greatly enhances the flexibility of design for the desired system output.

Physical Implementation of Compensators

The synthesis methods described earlier yield the transfer function required for effective compensation. Depending on the design method, either discrete compensator or continuous compensator transfer functions are obtained. The former is a rational function in z and the latter is rational in s. Since physical realization for continuous compensators are extensively covered in network theory

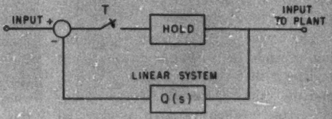


Fig. 10. Operational-instruction method of realizing a discrete compensator

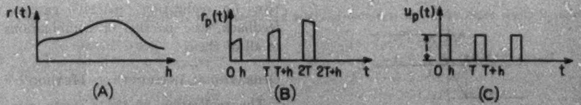


Fig. 11. Typical input to finite-pulsed system (A); sampler output in time (B); string of unit pulses (C)

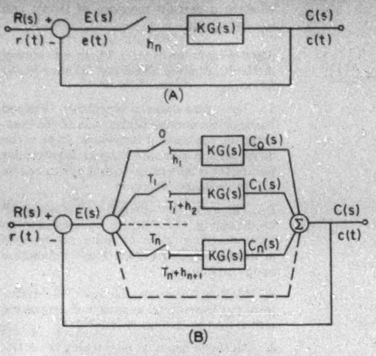


Fig. 12. Finite-pulsed feedback system

A—Typical system
B—Mathematical representation of (A)

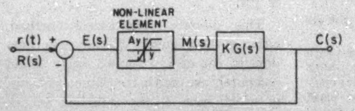


Fig. 13. Nonlinear feedback system

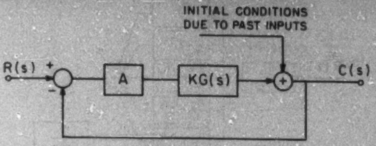


Fig. 15. Linear region of Fig. 13

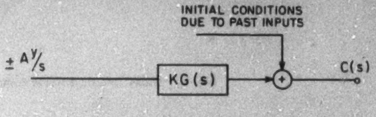


Fig. 16. Saturating region of Fig. 13

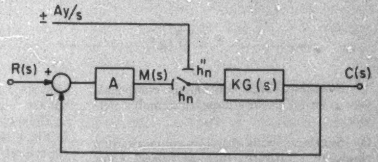


Fig. 17. Finite-pulsed feedback system of Fig. 13

their discussion is not elaborated on in this summary.

Discrete compensators can be physically realized through several methods, namely, programing, delay elements, pulsed *R-C* networks, and computer operational-instruction methods.

PROGRAMING METHOD[31]

With the growing field of digital computers and the development of various components in this area, the control engineer has the particular advantage of the use of such computers in the realization of discrete compensators. There are mainly three methods of programing: direct, cascade, and parallel, as well as cascade-paralleled programing. A typical diagram of a direct programing unit is shown in Figs. 7(A) and (B).

The choice of the method of programing depends on the number of storage registers, the computing time required, and the value of accumulation of round-off errors. Much research is directed toward construction of special computers for control systems.

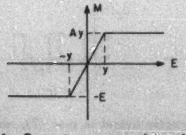


Fig. 14. Saturation type of nonlinearity

DELAY ELEMENTS METHOD[24]

This method indicates that the discrete transfer function where the numerator and denominator are rational polynomials in *z*, with the denominator of equal or higher order than the numerator, can be realized with the use of pure delay elements, voltage dividers, two algebraic summing devices, and two amplifiers. A typical configuration for such a realization is shown in Fig. 8. Methods of realizing pure delay elements are discussed in the literature.

PULSED R-C NETWORKS PROCEDURE[32]

This method utilizes *R-C* networks where the output is sampled in synchronism with its sampled input, as shown in Fig. 9. This performance is similar to that of digital computers, but with a considerable reduction in cost. However, this reduction is paid for with lack of flexibility. This procedure offers alternatives for realization; namely, cascade, feedback, and parallel or combinations of all of them.

OPERATIONAL-INSTRUCTION METHOD[11,25]

The realization of the discrete compensator using the pulsed *R-C* network requires in most cases a sampler and a zero-order hold in its output. This operational method removes this restriction by using a feedback realization procedure as shown in Fig. 10. This method can also be used for closely simulating the discrete compensator on an analog computer.

Sampled-Data Systems with Finite Pulse Width[10]

In the analysis of sampled data as discussed previously it is assumed that the sampler acts as an impulse modulator which is followed by a hold circuit before the information is fed into the plant. Based on this assumption the *z*-transform and the modified *z*-transform are readily applicable and have been successfully used.

In the situation where the pulses are of finite duration of comparable value with the system time constants, the *z*-transform does not yield the correct behavior for such systems. In such cases, a new form of Laplace transform called the *p*-transform could be used.

Consider an input to a sampler as shown in Fig. 11(A), the sampled output can be represented by Fig. 11(B), assuming a uniform pulse width *h* and fixed sampling period. Essentially, the sampler acts as a pulse modulator which can be mathematically described as:

Fig. 18. Computing facility, Donner analog computer model 30, and associated equipment

$$r_p(t)=r(t)u_p(t) \qquad (2)$$

where $u_p(t)$ is a unit pulse function shown in Fig. 11(C).

Taking the Laplace transform of equation 2 and effecting the complex integration in the v-plane yields for the transform of the output.[13] (The v-plane is a complex variable plane in which equation 3 is evaluated using the contour Γ. In this integration the variable s in $R(s)=r(t)$ is replaced by v, and in $U_p(t)$ it is replaced by $s-v$. For a complete discussion on the evaluation of the integrals in equations 3 and 7, the reader is referred to section 9.1, pp. 313–317 of reference 35.)

$$R_p(s)=\mathcal{L}[r_p(t)]=\frac{1}{2\pi j}\int_\Gamma R(v)\times \frac{1-\epsilon^{-h(s-v)}}{(s-v)[1-\epsilon^{-T(s-v)}]} \qquad (3)$$

where $R_p(s)$ is called the p-transform of $R(s)$ or $r(t)$ and denoted by:

$$R_p(s)=P[R(s)]$$

Evaluating the integral in the left half-plane enclosing only the poles of $R(v)$, $R_p(s)$ becomes:[13]

$$R_p(s)=\sum_{\substack{\text{roots of}\\ B(v)}} \text{Residue of} \left[\frac{A(v)}{B(v)}\frac{1-\epsilon^{-h(s-v)}}{(s-v)(1-\epsilon^{-T(s-v)})}\right] \qquad (4)$$

where

$$R(s)=\frac{A(s)}{B(s)} \qquad (5)$$

Equation 4 indicates that the p-transform of $R(s)$ is composed of rational functions of ϵ^{-hs}, ϵ^{-Ts}, and s. Once the transform of the sampler output is found, the usual Laplace transform procedure can be used to find the transform output $C(s)$.

To obtain the time output the inverse p-transform is effected which is based on the inverse modified z-transform procedure to give:

$$c(t)=P^{-1}[C(s)]=Z_m^{-1}[C^*(z,m)] \qquad (6)$$

where $C^*(z,m)$ is the modified z-transform of $C(s)$.

Alternately, evaluating the integral for $R_p(s)$ in the right half-plane of the v-plane[13] gives the following:

$$R_p(s)=\sum_{\substack{k=-\infty\\ k\neq 0}}^{k=\infty} R(s-jk\omega_r)\times \frac{1-\epsilon^{-jhk\omega_r}}{k}+R(s)\frac{h\omega_r}{2\pi} \qquad (7)$$

where

$$\omega_r=\frac{2\pi}{T}$$

Equation 7 yields for the limiting cases:

$$h=T,\ R_p(s)=R(s) \qquad (8)$$

and when $h=0$ and the pulse function approaches a unit impulse, $R_p(s)$ becomes:

$$R_p(s)=R^*(s)=\frac{1}{T}\sum_{k=-\infty}^{k=\infty} R(s+jk\omega_r) \qquad (9)$$

The form of $R_p(s)$ given in equation 7 can be used for approximate analysis of finite-pulsed sampled data in the frequency domain.

Approximation methods for finite-pulsed sampled data based on flat-top pulses has been used by Zypkin. Also for these approximations the modified z-transform can be applied as also discussed in detail by Zypkin.[15]

Finite-Pulsed Feedback Systems[33]

Extension of the techniques developed in connection with the p-transform, for exact analysis of feedback systems taking into consideration the actual pulse shape, can be easily applied. The mathematical details are straightforward and discussed by Farmanfarma;[33] however, the principle of the analysis is based on representing a finite-pulse feedback system, shown in Fig. 12(A), by its equivalent model shown in Fig. 12(B). Stability studies can be made, and the actual response can be easily calculated.

Application of Finite-Pulsed Systems Theory to Nonlinear Systems

The theory developed in connection with finite-pulsed feedback systems can be advantageously applied to nonlinear

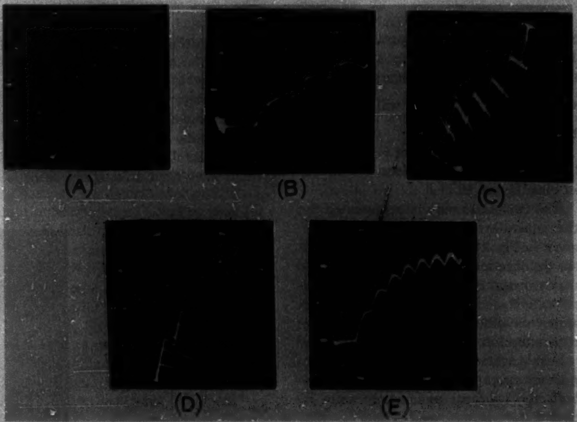


Fig. 19. Experimental study of finite-pulsed sampled-data systems where pulse width is varied

A—T=0.5 second
h=0.017, 0.2 second
B—T=0.5 second
h=0.017, 0.09 second
C—T=2.5 second
h=0.017, 1.0 second
D—T=0.5 second
h=0.3, 0.25, 0.2, 0.1 second
E—T=0.5 second
h=0.017, 0.2 second

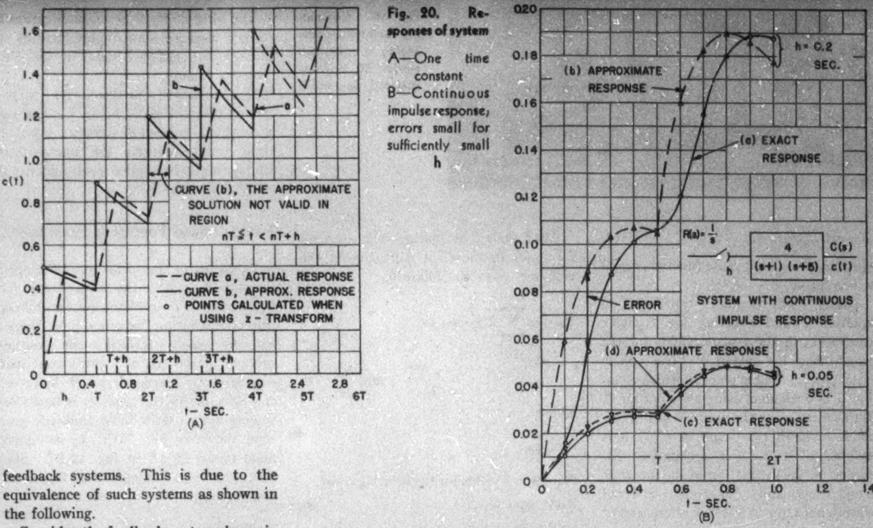


Fig. 20. Responses of system

A—One time constant

B—Continuous impulse response; errors small for sufficiently small h

feedback systems. This is due to the equivalence of such systems as shown in the following.

Consider the feedback system shown in Fig. 13 with a saturating-type nonlinearity indicated in Fig. 14.

The characteristics of this nonlinearity are:

$$M = A\,E \text{ for } -Y<E<Y$$
$$M = A\,Y \text{ for } \begin{matrix} E>Y \\ E<-Y \end{matrix} \qquad (10)$$

Such a system has two operating states: 1. when $-Y<E<Y$, the system will operate in the linear region; and 2. when $\begin{matrix} E>Y, \\ E<-Y \end{matrix}$ the system will operate in the saturated region.

The equivalent system of Fig. 13 in the linear and saturated regions are shown in Figs. 15 and 16.

The two states may be represented by the equivalent pulsed feedback system given in Fig. 17. In Fig. 17, h_n' and h_n'' are the durations for which the system remains in the linear and saturated states respectively. The switching from one state to another is achieved in zero time by assumption.

Such a system may be analyzed by the p-transform procedure.[33] In this method both the transient and steady state is obtained.

If the following steady-state condition is satisfied,

$$h_n' = h_{n+1}' = h_{n+2}'$$
$$h_n'' = h_{n+1}'' = h_{n+2}'' \qquad (11)$$

then the system can be considered to be composed of two periodic pulsed feedback systems for which the analysis techniques could be much simplified.

Pulse-Modulated Feedback Systems[34]

The effect of varying the pulse width as a function of the error is a useful means in compensating a continuous or pulsed system. Theoretically, this effect can be studied by a step-by-step procedure. This procedure yields information on the transient response for the first few sampling intervals.

Experimentally, variable pulse-width systems have been investigated in detail and typical useful results stabilizing a continuous system are indicated in the following section on experimental investigation. Further, theoretical and experimental investigations are indeed warranted and recent advances made in the theory of nonlinear difference equations and sampled-data systems can be advantageously applied to the study of pulse-width modulated systems.

Experimental Investigation of Sampled-Data Systems

The experimental results presented herein can be classified into two categories:

Vol.
Pt. 1
195

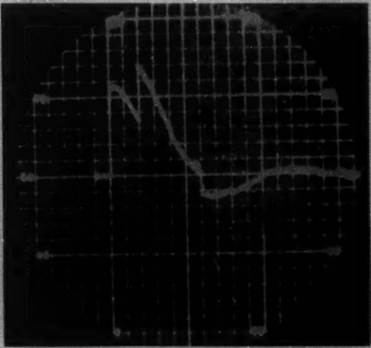
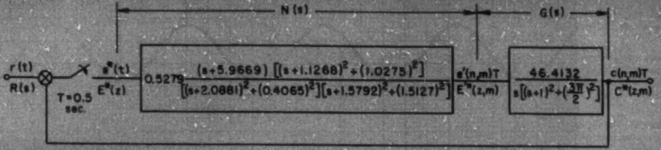


Fig. 21. Plant-forcing function e'(n, m)T for system in Fig. 2

Fig. 22. Sampled-data feedback system with fourth-order network compensation


N(s)
G(s)
r(t)
R(s)
T=0.5 sec.
e*(t)
E*(z)
0.5279 (s+5.9669) [(s+1.1268)² + (1.0275)²] / [(s+2.0881)² + (0.4065)²][(s+1.5792)² + (1.5127)²]
e'(n,m)T
E'*(z,m)
46.4132 / s[(s+1)² + (3π/2)²]
c(n,m)T
C*(z,m)

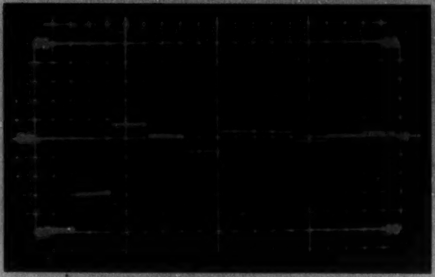
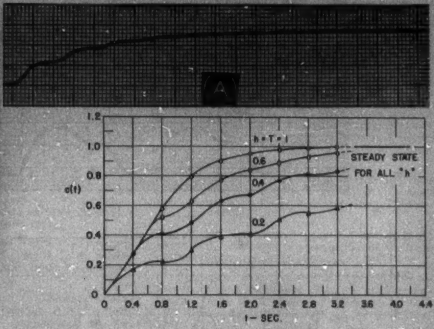


Fig. 23. Typical shape of error following hold in sampled-data feedback system



(B)

Fig. 24. Second-order pulsed feedback systems

A—Step response

B—Response of system to unit step input for various values of pulse width h

1. those results which are concerned with the checking of the theory and the effect of the simplifying assumptions made in the analysis; and 2. those results concerning the behavior of sampled-data systems under certain conditions such as variable pulse width.

These experimental results are only typical experiments made of sampled-data systems and by no means constitute an exhaustive study. These experiments were performed at the University of California in Berkeley on a specially built analog computer with special adjustments for sampling, holding, and other facilities for variable pulse width and/or sampling rate. This computer is shown in Fig. 18.

In the first category, Fig. 19 shows the actual response of a single integrator and two-time-constant system (continuous impulsive response) for values of pulse width h. These experimental curves can be compared with the theoretically computed curves [shown in Figs. 20(A) and (B)] based on the assumption of replacing the finite pulses by impulses and hence the use of the z-transform and the modified z-transform.

In the synthesis part and especially in the time-domain synthesis of sampled data using continuous networks, two typical figures are indicated.

Fig. 21 indicates the form of the plant forcing function $e'(n,m)T$ produced by the fourth-order network shown in Fig. 22. This plant-forcing function can be changed to achieve different responses depending on the complexity of the designed continuous network.

For the discrete compensator method of synthesis, Fig. 23 indicates the staircase step-shape input to the plant following the hold circuit of the discrete compensator.

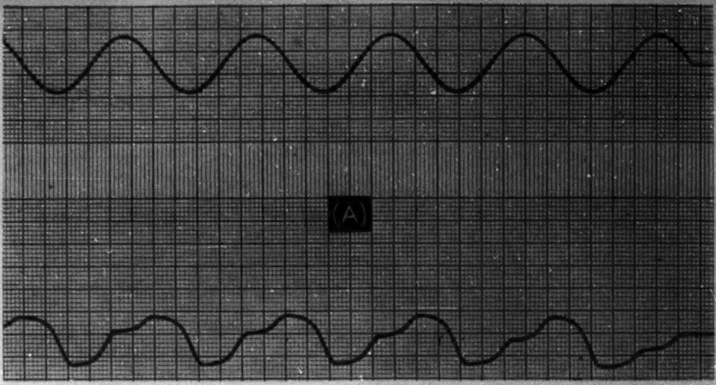
For the synthesis procedure based on the finite-pulsed theory developed for constant pulse width and sampling rate, Fig. 24(A) indicates the step response of a second-order pulsed-feedback system to be compared with the theoretically computed response of Fig. 24(B). Similar comparison for a sinusoidal input are indicated in Figs. 25(A) and (B). It is of interest to note that the response is not sinusoidal as in the cases of sampled-data systems (with almost zero pulse width) and continuous feedback systems.

Among the experimental results of the second category, Fig. 26 indicates the step response of the continuous system and pulse-modulated system. It is noticed that in both cases the pulse-modulated system has a stabilizing effect for it reduces the oscillations. Thus this system can be used as a compensating procedure for a continuous or sampled system provided the pulse width can be varied. Extension of this experimental study for higher-order systems both linear and nonlinear is undoubtedly of importance.

Conclusions

From this summary of some of the progress made in the field of sampled-data control systems, it appears that the linear theory is well organized and integrated. All the information related to linear systems is readily obtainable through any of the several methods of analysis and design mentioned in this progress note.[84] Other features of progress in this field have not been indicated because of limitations in space. Some of the problems which need more investigation in the coming years and which are presently undergoing extensive research are classified in the following three categories:

1. Nonlinear sampled-data control systems. This includes methods of analysis, design, and implementation of nonlinear digital controllers. Research on nonlinear difference equations and difference differential equations is necessarily warranted. The describing function method and the phase plane have been applied with a certain degree



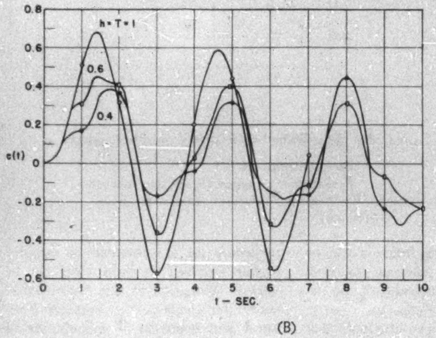


Fig. 25. Second-order system

A—Sinusoidal response
B—Response of system to sinusoidal input

$$R(s)=\frac{2}{s^2+4}$$

of success, but much ought to be done in the general case of nonlinear sampled-data systems.

2. Statistical theory of sampled-data systems. Extension of the Wiener Theory to sampled data has been successfully applied and the use of digital-filter theory in analysis and synthesis needs more research.

3. Adaptive sampled-data control systems. This category includes time-varying sampled-data systems which include varying sampling rate and pulse width or both, also the time-varying digital compensator element to adapt its operation to wide varieties of operating conditions and predicted signals operating in the system. Information theory fields can indeed be applied and extended to perfect the operation of sampled-data systems.

It is hoped that the material of this short discussion will summarize one aspect of the progress made in particular in linear sampled-data feedback systems, with much to be reported on the progress in other facets of this increasingly important field.

Nomenclature in Sampled-Data Field

$KG(s)$ = system-transfer function
s = complex variable of the Laplace transform
z = complex variable of the z-transform (equal to ϵ^{Ts})
w = complex variable in the ω-plane
$\sqrt{\;\;}$ = dummy complex variable
α = staleness factor always between zero and unity
k and n = positive integer
m = continuous variable always $0 \leq m \leq 1$
t = continuous variable, time
ω_r = sampling frequency
h = pulse width
T = period of sampler
$U_p(t)$ = unit pulse function
$r_p(t)$ = sampler output with a finite pulse width
$r^*(t)$ = sampler output with zero pulse width
$R_p(s)$ = Laplace transform of $r_p(t)$
$R^*(s)$ = Laplace transform of $r^*(t)$
$R(s)$ = Laplace transform of $r(t)$
$D^*(z)$ = discrete compensator transfer function
$C^*(z)$ = z-transform of $C(s)$
$C^*(z,m)$ = modified z-transform of $C(s)$
$C_n(s)$ = Laplace transform of the component of the output due to the $(n+1)$th input pulse
$P[\;]$ = p-transform of $[\;]$
$P^{-1}[\;]$ = inverse p-transform of $[\;]$
$z_m^{-1}[\;]$ = inverse modified z-transform of $[\;]$

References

1. CLASSIFIED BIBLIOGRAPHY ON FEEDBACK CONTROL SYSTEMS, PART I: SAMPLED-DATA SYSTEMS, **Thomas J. Higgins, Robert W. Greer.** *AIEE CP 58-1269.*

2. A SELECTIVE BIBLIOGRAPHY ON SAMPLED DATA SYSTEMS, **Peter R. Stromer.** *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, New York, N. Y., Dec. 1958.

3. THE DYNAMIC OF AUTOMATIC CONTROLS, **R. C. Oldenbourg, H. Sartorious; H. L. Mason,** translator. American Society of Mechanical Engineers, New York, N. Y., Feb. 1948, p. 276.

4. A NOTE ON THE OPERATIONAL SOLUTION OF LINEAR DIFFERENCE EQUATIONS, **E. I. Jury, F. J. Mullin.** *Report no. 60*, Electronic Research Laboratory, University of California, Berkeley, Calif., Issue 199; also *Journal*, Franklin Institute, Philadelphia, Pa., vol. 266, no. 3, Sept. 1958, p. 189.

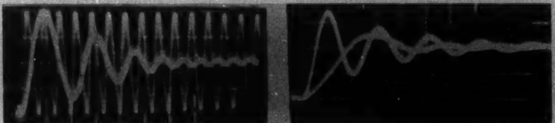


Fig. 26. Step responses of second-order system for continuous system and for pulse-width modulated system

5. General Synthesis Procedure for Computer Control of Single and Multiloop Linear Systems, **R. E. Kalman, J. E. Bertram.** "Proceedings of the Computers in Control Systems Conference," *AIEE Technical Publication T-101*, May 1958, pp. 130–36.

6. The Design of Sampled-Data Feedback Systems, **Gladwyn V. Lago, John G. Truxal.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 73, Nov. 1954, pp. 247–53.

7. Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation, **William K. Linvill.** *Ibid.*, vol. 70, pt. II, 1951, pp. 1779–88.

8. Signal Flow Reductions in Sampled-Data Systems, **J. M. Salzer.** *Wescon Convention Record*, Institute of Radio Engineers, pt. 4, 1957, pp. 166–70.

9. Input-Output Relationships of Multisampled-Loop Systems, **George G. Lendaris.** *Report no. 214*, Electronic Research Laboratory, University of California, Issue 60, Sept. 23, 1958.

10. The Analysis of Sampled-Data Systems, **J. R. Ragazzini, L. A. Zadeh.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 71, Nov. 1952, pp. 225–34.

11. The Pulse Transfer Function and Its Application to Sampling Servo Systems, **R. H. Barker.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. IV, Dec. 1952, pp. 302–17.

12. Analysis and Synthesis of Sampled-Data Control Systems, **Eliahu I. Jury.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 73, Sept. 1954, pp. 332–46.

13. Synthesis and Critical Study of Sampled-Data Control Systems, **Eliahu I. Jury.** *Ibid.*, pt. II (*Applications and Industry*), vol. 75, July 1956, pp. 141–51.

14. Addition to the Modified z-Transform Method, **E. I. Jury.** *Wescon Convention Record*, Institute of Radio Engineers, pt. IV, Aug. 1957, pp. 136–56.

15. Differenzenglichungen der Impuls-Und Regeltechnik (book), **J. S. Zypkin.** Verlag Technik, Berlin, Germany, 1956.

16. Compensation of an Error-Sampled System by a Multirate Controller, **George M. Kranc.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 76, July 1957, pp. 149–59.

17. The Analysis of Sampled-Data Control Systems with a Periodically Time Varying Sampling Rate, **E. I. Jury, F. J. Mullin.** *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, vol. PGAC-4, no. 1, May 1959.

18. Analysis of Cyclic-Rate Sampled-Data Feedback-Control Systems, **R. E. Hufnagel.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 77, Nov. 1958, pp. 421–25.

19. Analysis of Linear Sampled-Data Systems with Finite Pulse Width: Open Loop, **G. Farmanfarma.** *Ibid.*, pt. I (*Communication and Electronics*), vol. 75, 1956 (Jan. 1957 section), pp. 808–19.

20. Network Compensation of Error-Sampled Feedback Control Systems, **J. Sklansky.** *Technical Report T-7/B*, Electronic Research Laboratories, Columbia University, New York, N. Y., Apr. 1, 1955.

21. Time-Domain Synthesis of Sampled-Data Control Systems, **E. I. Jury, F. W. Semelka.** *Transactions*, American Society of Mechanical Engineers, New York, N. Y., Nov. 1958.

22. Analysis of Control Systems Involving Digital Computers, **W. K. Linvill, J. M. Salzer.** *Proceedings*, Institute of Radio Engineers, vol. 41, July 1953, pp. 901–06.

23. Transient Analysis of Sampled-Data Control Systems, **G. W. Johnson, D. P. Lindorff.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 73, July 1954, pp. 147–53.

24. Discrete Compensation of Sampled-Data and Continuous Control Systems, **Eliahu I. Jury, William Schroeder.** *Ibid.*, vol. 75, 1956 (Jan. 1957 section), pp. 317–25.

25. Characteristics of Sampling Servo Systems in Automatic and Manual Control (book) **C. Holt Smith, D. F. Lawden, A. E. Bailey; A. Tuston,** editor. Butterworth's Scientific Publications, London, England, 1952.

26. Sampled-Data Processing Techniques for Feedback Control Systems, **Arthur R. Bergen, John R. Ragazzini.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 73, Nov. 1954, pp. 236–47.

27. Uber die Synthese Von Impuls Systemeu der Automatischen Regelung und Steuerung, Fachtaging Regelungstechnik, **J. S. Zypkin.** Heidelberg University, Heidelberg, Germany, 1956.

28. Minimization of Total Square Error in Sampled-Data Control Systems, **B. Bharucha.** *M. S. Thesis*, University of California, June 1958.

29. Correlation Between Root-Locus and Transient Response of Sampled-Data Control Systems, **Eliahu I. Jury.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 74, 1955 (Jan. 1956 section), pp. 427–35.

30. Root-Locus Method of Pulse Transfer Function for Sampled-Data Control Systems **Mashiro Mori.** *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, PGAC-3, Nov. 1957.

31. Frequency Analysis of Digital Computers Operating in Real Time, **J. M. Salzer.** *Proceedings*, Institute of Radio Engineers, vol. 42, Feb. 1954, pp. 457–66.

32. Pulsed RC Networks for Sampled-Data Systems, **J. Sklansky.** *Convention Record*, Institute of Radio Engineers, pt. II, 1956, pp. 81–99.

33. General Analysis and Stability Study of Finite Pulsed Feedback Systems, **G. Farmanfarma.** *Series no. 69*, Electronic Research Laboratory, University of California, Issue no. 198, Nov. 1957.

34. Experimental Study of Pulse-Width Modulated Feedback Control Systems, **Noor Billawala.** *M. S. Thesis*, University of California, Jan. 1958.

35. Sampled-Data Control Systems (book), **E. I. Jury.** John Wiley & Sons, Inc., New York, N. Y., Nov. 1958.

Considerations Pertaining to the Design of an Electronic Telephone Switching System

B. BRIGHTMAN
MEMBER AIEE

M. P. TUBINIS
NONMEMBER AIEE

THE U. S. Army Signal Corps required a basic telephone switching system which could fulfill the functions of a local central office and provide the basis for a long distance (LD) switching center. In addition, a complete switching network was required, built up of a series of local and LD switchboards.

Due to the tactical character of this network it was required that each office be self-contained and fully transportable either by air or ground. Therefore, weight and size was of the utmost importance. At the same time, however, in order to provide maximum service under emergency tactical conditions the LD central offices were to provide nonblocking facilities. The trunk capacity of each LD office was specified at 160 trunks and 80 links. The local central office, however, was not required to be of a nonblocking type, but did require a relatively high grade of service comparable to a commercial system. Two-hundred-line 80-trunk 60-link circuits were specified for this board. The conflict of requirements here, that is, weight and size versus line, trunk, and link capacity, brought forward the request for an all-electronic system utilizing solid-state circuitry to the maximum extent.

To provide for ease of maintenance, standardization of circuitry in the two types of switchboards was essential. This principle, therefore, underlined the basis for the design of the switchboards.

Skeleton models were built and performed satisfactorily. The first two fully equipped service test models have been produced and are now undergoing production tests.

Nomenclature

TDM = time-division multiplex
SD = seize detector
LC = link circuit
CC = control connector
SSD = seize-signal detector
RA = register allotter

Paper 59-182, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 24, 1958; made available for printing April 8, 1959.

B. Brightman and M. P. Tubinis are with the Stromberg-Carlson Company, Rochester, N. Y.

The authors gratefully acknowledge assistance provided by members of the Switching Branch, Data Processing Division, Communications Department of the U. S. Armed Services Research and Development Laboratories, particularly G. Bartle and T. Pfeiffer. Acknowledgments are also extended to the numerous engineers of the Telecommunication Division of Stromberg-Carlson, in particular to S. Weiner, J. G. Pearce, G. O. K. Schneider, and T. Frankel for the contributions made in solving the many problems that existed.

SR =subscriber register
RF =register finder
LA =link allotter
RCF =recall finder
PR =position register
MDR =master-clock generator
SDC =seize-detector control
RS =route selector
MSD =multifrequency signal detector

Choice of Cross Pointing

A summary of various types of speech cross pointing soon showed that in order to meet the space and weight requirements full utilization of the high-speed characteristics of transistors would be required. For example, a space-division method of providing a speech connection in a balanced nonblocking system would require a cross point per line or trunk per connection to every other line or trunk in the board. For a 160-trunk board this would require (160^2-160) 2 to give 50,880 transistors. However, by utilizing a time-division system the number of transistors required could be reduced, depending upon the trunking configuration, to 4,480 transistors to provide exactly the same facilities. The overwhelming space and weight advantage clearly indicated that our design should be based on a time-division system. After a feasibility study, our design of the complete system was started, and the entire project was split into three main fields of endeavor. These were:

1. Design of a TDM speech path.
2. Design of signaling equipment (such as voice-frequency receivers).
3. Design of an over-all switching system based on items *1* and *2*.

Separate papers[1] are being presented on items *1* and *2* and will not, therefore, be discussed at this time, except as will be necessary to an understanding of this paper.

Trunking

The first tentative system design was based on the following premises.

To meet the nonblocking requirements a link circuit to connect a calling line to a called line would have to be provided on the basis of 1/2 (total line and trunk capacity) of the board. Provision of this very large number of links could be achieved by two methods. One method was to assign a fixed time position per link which required that a line or trunk circuit had to be capable of selecting any one of the total time positions or slots available in order to be able to set up a speech connection. The other method was to allot a fixed time slot to a line circuit and make each link circuit capable of switching to any one of these allotted time slots. This latter method was finally adopted after consideration of the fact that it was obviously desirable to minimize the amount of equipment on a line-circuit basis. It was decided, therefore, to make the link circuit the variable factor as regards time-slot storage and also for it to provide access to supervisory tones such as ring tone, etc.

Before any calculations could be made regarding equipment required, it was obviously essential to know the total number of time slots that could be used within a given frame period, in order to reconstruct the original information after it had been sampled. It is common knowledge concerning time-division multiplexing that the information sampling rate should be at least equal to two times the maximum frequency being transmitted.

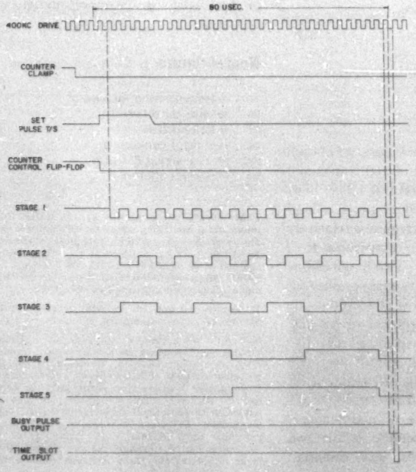


Fig. 1. Timing diagram of time-slot storage counter

At this point it is well to stop to consider the specifications relating to the transmission characteristics of the system, as the choosing of a suitable frame time, sampling time, guard time, etc., was based on these specifications. The salient requirements are as follows:

1. The transmission path had to be capable of carrying data information which limited the maximum delay distortion that could be tolerated through each switchboard to 10 μsec (microseconds).
2. Adjacent channel crosstalk had to be greater than 50 db (decibels) below 0 dbm (db referred to 1 milliwatt) reference level.
3. The insertion loss of the board had to be limited to 0 db ± 0.5 db.

Calculations showed that a frame repetition rate of 12.5 kc would enable us to meet the requirement in item *1* with a relatively small amount of phase equalization circuitry.

In order to meet the requirement of item *2* we had to be sure that the time between successive sampling periods was long enough for the energy stored in the highway (due to stray capacity effects) to decay. This could be met by providing a sufficiently long guard time. However, to ensure that this crosstalk limit was met, irrespective of highway loading and length, it was decided to include an artificial discharge circuit to dissipate any signal energy still present after the assigned sampling period.

This left the insertion loss specification mentioned in item *3* as a factor to be considered in deciding what sampling-pulse width would be used. The use of resonant transfer-type time division is dependent upon pulse-width stability to ensure full transference of sampling power. Therefore, any variation in rise or fall times of two connected multiplexed gates would give an equivalent percentage variation in transferred power level. Due to this, the wider the sampling pulse, the smaller the effect of differences between rise times of differing transistors.

With these factors in mind, the selection of sampling, guard, and frame times was made, bearing in mind the effects they would have on the switching and storage circuitry.

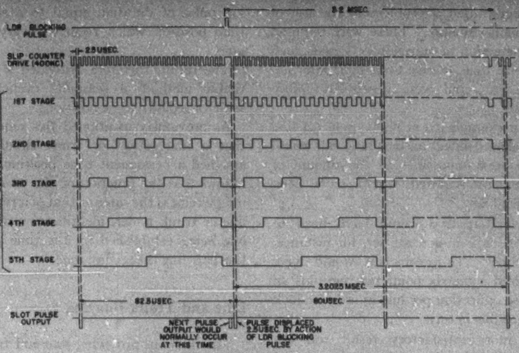


Fig. 2. Timing diagram of slip counter

For simplicity of circuitry and the use of as low a number of components as possible, it was decided that binary storage of time-slot information would be used. A binary number of time slots per frame was therefore desirable. As a result, a frame period made up of 32 time slots was investigated, which, by using a one-to-one ratio of sampling-to-guard time, would provide directly the basic clock frequency of the system. The clock frequency, however, had to be such that transistor characteristics would enable reliable working of the switching circuits, while the sampling time of each time slot had to be sufficiently long to enable the insertion-loss specification to be met. As a result, it was decided that 1.25-μsec sampling and 1.25-μsec spacing between sampling pulses would meet transmission requirements. This gave an over-all pulse cycle of 2.5 μsec per time position or a clock frequency of 400 kc. As already stated, 32 time slots per frame were required; therefore, the frame time was given as (2.5 μsec) 32 which gave 80 μsec per frame or $10^6/80$ which equalled 12.5 kc. With these characteristics specified, work on the transmission and switching circuitry was started.

Storage of Time-Slot Information

The next development stage provided for the manipulation of time-slot intelligence within the switching circuitry. This refers to the manner in which any two terminations could be connected via one of the available number of time periods.

One method, for example, to connect two lines is to allot a common time period to each of the terminations. This, however, requires storage of the time information in each of the two lines.

Two types of storage are available: static and circulating. A look at the static-type memory soon indicated that a large number of components would be required as a separate store per line per time slot would be required together with associated input and output gating. When it is considered that a store such as this would have to be provided per termination, then, for a nonblocking board, a very large number of components would be required. Very rapidly, therefore, the conclusion was reached that a circulating type of memory would be used.

The selection of a transistor counter-type store to provide this function was made, based on the premise that as the system would be of a common control type, fail-safe type of circuitry would be used as far as possible. A transistor-type counter store more nearly approached this requirement. For example, if a failure of the common pulse-drive supply occurred, then each time store in operation at that time would maintain its position with relation to each other store in use; so that when the drive supply was re-established the time relationship between time stores would be maintained, which would not be the case if a delay line had been used. Consideration was also given a slaved multivibrator-type circuit but was discarded for the same reason as for the delay line.

The transistor binary counter that was developed for time memory met the requirements of size, simplicity of operation, and ease of manipulation. The use of a binary counter in this application as opposed to a decimal-type ring counter had several advantages and yet did not have the disadvantage of requiring complicated multiple-output gates in order to extract each count point. The main advantage of using the binary counter was obviously size, as we required at least 32 separate time positions in a frame. With a ring counter-type store, 32 flip-flops would have been required or a 4 by 8 matrix-type count, as against a 5-stage flip-flop binary counter. Frequency breakdown was another desirable feature of the binary counter not given by a ring counter, as only the first stage of the binary is a high-speed count and each subsequent stage is frequency halved. The usual disadvantage of a binary-type store which requires multiple-output gates did not arise, as, in the type of operation in which it was required, we were only interested in obtaining a single output. True, this output had to be capable of producing a pulse related to any of the 32 time positions in the frame but this was obtained by initially triggering the counter at the time interval required so that its output pulse was directly related to the time period of the count initiating pulse; see Fig. 1. For example, if 32 separate counters are considered, each in a static condition, and reset to the same position and each then triggered consecutively by one of the 32 time-position pulses, then at the end of the frame each of the 32 counters would be running but each would be out of step by the count of one compared with the preceding counter. The over-all result therefore, if each of the outputs of the 32 counters were added timewise, would be to reconstruct the originating 32 separate time positions. Each counter therefore would provide an output pulse in phase with the time position of the pulse which initially triggered it.

Basically, therefore, all that is required to connect two lines together out of a multiple via a coincident time position, is a binary counter in each of the two line circuits preset to a fixed position. The trigger input to the two counters would be gated in simultaneously with a pulse whose time position is related to a "free" pulse within the 32 time positions constituting a frame. The two counters in each of the lines would, therefore, each produce a pulse coincident with each other which if applied to a multiplex switch would provide a speech connection between the lines.

Scanning of Voice-Frequency Detector Circuits on a TDM Basis

The various voice-frequency detectors that would be required throughout the system caused a problem in the use of common circuitry. These were required for detection of supervisory signals such as seize tone, release tone, digit information, etc., and were necessarily bulky in size.

For example, to provide an SD per line in order to detect an incoming call would have been impossible, as the amount of equipment required would have been tremendous.

An alternative would have been to provide a common SD per, for instance, 100 lines. However, this would have required a matrix point per line plus an audio connection per line in order to look at each line in turn.

A more satisfactory result was desired, therefore a method was evolved for scanning on a TDM basis so that individual space-division access to each line circuit was not required.

In order to utilize the already existing TDM switches providing speech access from each line to the multiplex highway, it was very quickly realized that if each line circuit was permanently allotted a time position and continuously pulsed, regardless of whether the line was busy or free, then a subscriber requiring a connection would transmit seize tone, which would be gated onto the multiplex highway via the line-circuit multiplex switches. It can be seen, however, that as each line is allotted its own individual time position, no connection between any two lines can be established without going through an intermediate time-coincidence conversion circuit.

Therefore, if a common seize detector was provided which had access to the multiplex highway and also had the facility to look at each time slot in turn for a period depending upon the recognition time of the detector, then it would, in effect, provide facilities to scan each line circuit without requiring any space-division equipment

The solution of this problem produced an economical, simple, and reliable method. All that was basically required was to provide a time-slot storage counter with the ability to, in effect, "slip" a time position after a suitably long period of looking at that time position. This was done by producing a "blanking" pulse, to inhibit the clock-driving pulse of the time-slot store (Fig. 2). The result was that the output of the time store provided a pulse shifted in time for each application of the blanking pulse. For example, it was found that the recognition time of the SD was 3.2 ms (milliseconds), therefore, a count-down counter was designed to provide a blanking pulse of duration equal to a single time position which recurred every 3.2 μsec.

The result, therefore, provided a very simple method of scanning cyclicly a series of separate time positions.

As previously mentioned this required that a line circuit or trunk circuit be assigned a permanent time position. It follows that the function of the SD scanning circuit is the same as that of a typical line or trunk finder, identification of the line being established by the time position carrying the seizure tone.

Line and Trunk Busying

The mode of providing line and trunk busying was the next problem to be tackled. Similar to the trunk and line-finding problem just discussed, a space-division method of providing a static store such as a flip-flop per line was obvious but required a large amount of space-division circuitry and interconnections. An investigation was instigated into methods of providing a circulating memory to indicate busy or free conditions, which would result in a single-wire busy-test facility. For example, if it was possible to indicate on a single wire what time positions were already in use, then it would be a simple matter for the routing circuitry to interrogate this wire to ascertain which time positions were already busy before setting up a required connection.

This gave a simple and economical method of providing busy testing facilities and the circuitry required was already provided. It has already been established that any common circuitry connected to a calling line would be connected via a time position generated by its own time-slot store. Therefore, if the output of this store was connected to a common bus similarly connected to all other circuits that generated a time slot, then any of these circuits generating a time pulse would also feed that pulse onto the common bus. By comparing the presence or absence of this pulse referenced in time with the known time position of a required connection then a line-busy condition would be detected by the presence of the busy pulse inhibiting the connections. The problem here was to ensure that the busy-line pulse was already present before actual testing, as differences in rise times of gating circuitry providing the coincidence checking might result in a faulty test condition. This was overcome by making the busy pulse originated by a time store, equal in width to a full cycle of the main-drive frequency of 400 kc. This provided a busy-pulse width of 2.5 μsec as against the time-slot pulse width of 1.25 μsec, thereby ensuring that a busying pulse would always be present on the busy-test bus before a coincidence check with a time slot was made.

This common bus, therefore, provided two basic functions: 1. It inhibited the common SD from looking at a line once that line was busied. 2. It provided busying intelligence to the line selector to enable that circuit to determine if a terminating connection was already in use.

Speech Cross Pointing Between Two Line Circuits

Since it was established that the line circuits would be assigned permanent separate time positions, the problem of providing a coincidence point for a speech connection was solved by providing an intermediate switching point between any two lines. This is called the link circuit (LC) and has, in effect, two separate time stores. Each time store controls a multiplex switch which allows the calling and terminating time-slot information of the relevent line circuits to be read into the line circuit to provide access to each line. Conversion at this point into audio intelligence was made so that it could be resampled by the time position of the other line, and also to provide an effective connection between the two circuits.

An understanding of this basic thinking behind the system will enable the following system description to be more fully understood.

Systems Operation

To describe the system which has been developed, some preliminary background information will be required. As stated previously, it is a TDM system utilizing 32 time slots of 1.25 μsec duration at a repetition rate of 80 μsec for 12.5 kc. A guard period of 1.25-μsec duration which precedes and/or follows each sample pulse, is used to clamp the previous signal period to reduce crosstalk to the required levels. These pulses are generated by the use of a binary counter driven from a 400-kc oscillator. In the local switchboard 200 lines and 80 trunks have to be provided with individual identities. This is accomplished by assigning 28 pulses to 10 separate paths or highways. These 10 highways are physically separated from each other and can only be switched

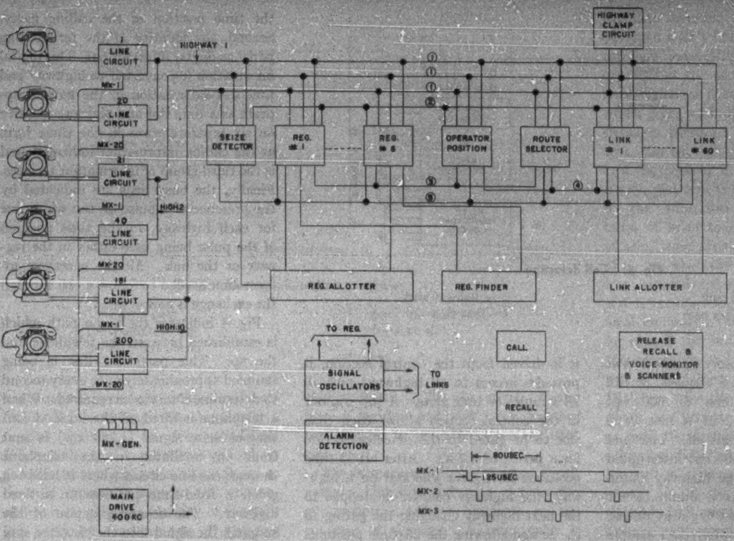


Fig. 3. Block diagram of local switchboard

1—Multiplex highways
2—Calling time slot and highways busses
3—Digit-transfer busses
4—Terminating time slot and highways busses
5—Busy-test busses

at the link circuit which will be described later. Since the 10 highways are isolated, the same pulses are used on each of the 10 highways. That is, pulse *1*, or *MX1*, will drive a send and a receive switch in a line circuit on each of the 10 highways simultaneously. The same is true for *MX2*, etc. In this way, 280 (10×28) individual identities are obtained.

The line circuit mentioned is the circuit at which the audio signal arriving is chopped up and distributed in time on to the highway. At the same time, any signal passing through the exchange in this time period is gated off the highway, converted to audio, and transmitted as an audio signal. All signals arriving and leaving the switchboard are audio signals. Each line circuit is assigned a permanent time slot, i.e., is constantly being sampled by a fixed *MX* pulse, on a given highway. This identity is unique and can be assigned a number which will select this line circuit. The method for selecting this number will be discussed later.

Up to this point, we have covered the subscribers' access to the exchange via line circuits. Now, the question of the amount and type of common equipment in the local switchboard which is used by the subscriber will be discussed (Fig. 3). The first item encountered is the SD which is a single unit consisting of a control connector (CC), seize-signal detector (SSD), and a control unit. These subunits will be covered with more detail when a call is followed through the exchange. The next unit utilized is the register allotter (RA), a single unit, which distributes the call into one of six subscriber registers (SR). Following the registers, is a register finder (RF), another single unit, whose function is to sequence the use of the route selector, which in itself is a single unit. The link allotter (LA), whose function is similar to the RA, distributes the call to any one of 60 links which are the only pieces of equipment remaining in the call during the entire conversation period.

In order to release the equipment a scanner periodically samples each link for release signal. This scanner also searches for recall, and is utilized for voice monitoring in order to release a link if it has been engaged and fails to release for some reason. Further discussion of this point will be carried out later. A ring-trip scanner also samples each link for the purpose of shutting off ringing.

The operator's position has direct access to every link on a common audio path that is controlled by switches. A call and an RCF, each an individual unit, seek out links which have been premarked as operator's calls or recalls, and provide both searching and cueing action for the operator. The operator's keyboard is of the throw-off type which can connect a call in a premarked link, extend a call, or give information after connecting to this link with the operator's key. However, after restoring the key to its normal position, the operator cannot reenter an established call unless a recall condition is placed on this link. A separate service link is provided for extension of calls originated by the operator. Digital information is keyed into a PR which is solely for the use of the operator.

The remaining equipments which make up the electronic section of the switchboard are the pulse and signal generating equipment, highway clamp circuits, routine test circuits, and alarm scanning and detection equipment. The binary counter which generates the *MX* pulses is denoted as the *MX* generator. The 400-kc generator, or what we refer to as the MDR generator, is the master clock which drives the *MX* generator and all other binary counters utilized in the exchange for generating sample pulses for driving audio switches similar to the line circuit. The signal generator and supervisory tone generator provide the audio tones for signaling in the system and any necessary supervisory tone. For signaling, 1,700 cps (cycles per second) is utilized as the seize signal; 1,900, 2,100, 2,300, 2,500, and 2,700 cps, utilized in pairs make up digital signals; 2,900, plus each of the previously mentioned frequen-

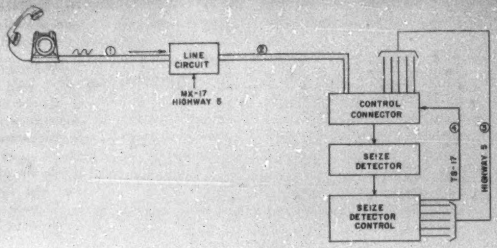


Fig. 4. Call detection

1—Audio path
2—Multiplex path
3—Highway mark
4—Time slot

cies, is used for supervisory signals such as release, recall, interlock request, or call terminated. Ring tone is 600 cps modulated at 20 cps, at a rate of 1 second on and 2 seconds off. Line and trunk busy tone are 600 cps interrupted at different rates. The highway clamp circuit simply acts as a dump circuit which is opened on every highway during the guard period following each sample pulse. Its purpose is to discharge stray capacities which would retain part of the previous conversation sample and allow this to leak into the next sample period. The routine test equipment is provided to allow service personnel to send calls into preselected registers and links for checking their operation. The alarm circuit is divided into two groups. One section is utilized to constantly monitor each highway. A constant tone of 1,000 cps is gated out in time position 32 and is detected by this circuit. If for some reason, this tone should not be detected, an alarm is given indicating a faulty highway. The second section is really a scanner which scans for various alarm conditions throughout the system, and upon finding one, gives a visual and audio indication.

This covers all the equipment other than the charger and batteries in the exchange which provide a 24-volt center-tapped output giving three potentials: plus, minus, and ground.

In tracing a call through the exchange certain preset conditions exist and certain scanning actions are in progress. The RA will have premarked a free register so that the next call coming into the exchange will be handled by this register. In a similar fashion, the LA will search out a free link and also premark this link to handle the next call.

The scanning action which is occurring is being carried out by the SDC. The CC is connected to all highways and since it is driven from the control section, it provides access to all highways and all 28 samples or time slots. A slip counter in the control section allows each time slot to be gated to the detector section for a period of 3.2 ms. After all 32 time positions have been checked on a highway, the highway counter is stepped to the next highway changing the gating to the CC and allowing the 28 time positions on this highway to be checked. This scanning continues until a seize signal (1,700 cps) is detected. At that time, the SSD sends a pulse to the control section which stops the scanning action of the unit and initiates a demand condition on the RA.

A little more information at this point will aid in following the call through the exchange. Three separate bits of information are transferred internally in the exchange. First, the identity of the calling party. This identity is in the form of highway and time-slot information. The highway is transferred by marking two of five leads and the time slot by gating out a negative pulse in the time position of the calling party. Second, the identity of the terminating party as digital information. This digital information is converted to highway and time-slot information in the RS and appears as a time slot and highway identity on six separate wires in the same form as the calling information indicated and is the third group of information accesses. Finally, the busy lines are indicated by the presence of a pulse on ten wires, one for each highway, in the time position of the pulse being used either in the register or the link. All the necessary information needed to follow a call through the exchange is now at hand.

Fig. 4 indicates the audio path which is established between the telephone and the SD. This particular line is being sampled approximately once every second to determine if service is required. When a telephone is lifted off the hook, a continuous seize signal (1,700 cps) is sent from an oscillator in the telephone through the line circuit where it has been given a fixed-time position on a fixed highway. The scanning action of the SD gates the signal into the detector stag where it produces a mark for the control section, stops the scanning action and thus retains the identity of the party originating the call. At this time a demand condition is placed on the RA to transfer the call to one of the registers. At this time, assuming that no other call exists in the exchange, no pulses are present on the busy-test busses. A busy pulse on one of these busses inhibits the SD from sampling this time position. However, the scanner does not advance, skipping a busy pulse in the scanning process.

Fig. 5 indicates the next step in establishing the call. As previously stated, a register is preallotted by the RA which

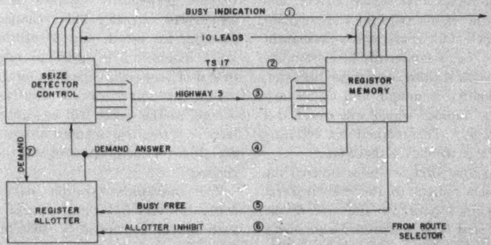


Fig. 5. Transfer of call to register

1—Busy-test busses
2—Time-slot transfer bus
3—Highway-transfer busses
4—Demand-answer mark
5—Register busy-free mark
6—Allotter-inhibit mark
7—Connect-demand mark

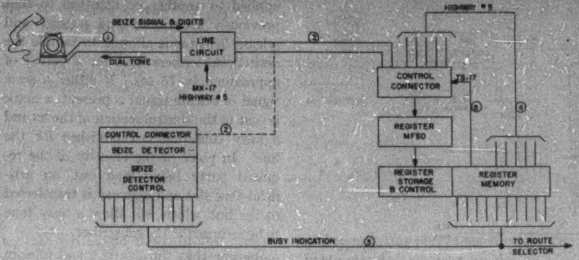


Fig. 6. Keying information to register

1—Audio path
2—Multiplex path
3—Busy-test busses
4—Highway marks
5—Time slot

has sequentially sampled individual busy-free leads from the registers, stopping on a register indicating a free condition. When a demand condition is placed on the allotter by the SD, a write signal or demand answer is sent to the SDC enabling it to write the identity of the calling party by marking two of the five highway busses between the SD and the register. At some time during this writing period, which would normally persist for 160 μsec, a time slot is transferred to the register which will open a gate allowing the 400-kc MDR pulse to drive a 5-stage binary counter, gating out a time slot to a CC identical to the one used in the SD. Simultaneous with the write signal to the SD, a read signal is sent to the preselected register allowing the identity of the highway to be stored in the static form, and enables the time slot from the SDC to trigger a binary counter for dynamic storage of the time position of this call. The register indicates busy after storing the information, showing this condition on the busy-free lead to the RA. After a write-read time of 160 μsec, the RA samples the busy-free lead from the register to determine if the information has been stored in the register. If, at this time, the register has failed to read this information and has not indicated busy, the allotter alarms this register and steps on to the next register, repeating the read-in process a second time. If a second attempt fails, the demand answer is removed resetting the SD; the call is abandoned, and a major alarm indicated.

Assuming that the register is operating satisfactorily, the information is read from the SD and a busy indication is forwarded to the allotter. The write signal is removed from the SD, resetting the SDC which proceeds to scan the exchange again. The RA steps off the busied register and seeks out another free register to premark it for the next call.

The memory unit in the register, which gates a highway and time slot to the CC, performs several functions in the progress of the call. A busy pulse is gated out of the register on to the busy-test bus associated with this highway in the time position of the calling party. This pulse inhibits the SDC from further sampling of this time position and is utilized by the RS in determining whether a line or trunk is busy or idle. At this time, the register acts as the terminating party (Fig. 6). The act of storing the identity of the calling party triggers the control section of the register. An MSD can be considered as the ear of the register while the control section can be considered as the vocal section. The initial action of the register is to check that the seize signal is true. No voice guard is present in the SD and a fairly wide band is sampled for a seize signal to obtain the high-speed response for the fast scanning action. The MSD is equipped with a full voice guard and narrow band-pass filters (band width ±50 cps around the signal frequency). The seize signal is passed into the detector which gives an indication if a true seize signal is present. If the signal is absent, the call is rejected after 20 ms and the register is reset. If the signal is real, dial tone is sent from the register to the subscriber following a timed burst of seize tone. This initial burst of seize tone preceding the dial tone is of no use when the call is placed directly from the subscriber; however, when the call is coming over a trunk from other local exchanges, this burst is responsible for closing the audio path of the link in the preceding exchange. This function will be explained in more detail later. Dial tone trips the seize tone in the telephone and persists for 10 to 20 seconds, or until the first digit is keyed. If the subscriber fails to key within this period, the call is rejected and the register is reset. A 10- to 20-second interval is allowed for every digit keyed until sufficient information is keyed into the register. When sufficient digital information has been stored, the input to the register is blocked and a demand condition placed on the RF.

Fig. 7 illustrates the next item of control equipments which are utilized in the progress of a call. The RF is brought into action by the demand from a register, and answers the demand by marking a lead to this particular register. This answer allows only the digital information to be read out of the register in 2-out-of-5 form on 15 wires. These wires are common to all registers, and as a result only one register is given access to the RS at any one time. The RF provides this controlling function. The appearance of digital information on the busses between the register and RS

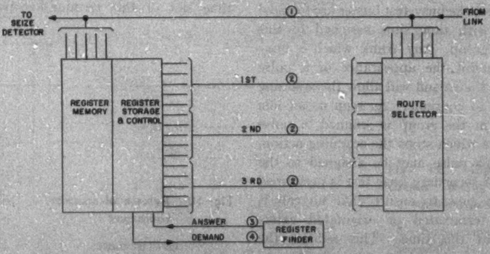


Fig. 7. Transfer of digital information from register to route selector

1—Busy-test busses
2—Digit-transfer busses
3—Demand-answer mark
4—Connect-demand mark

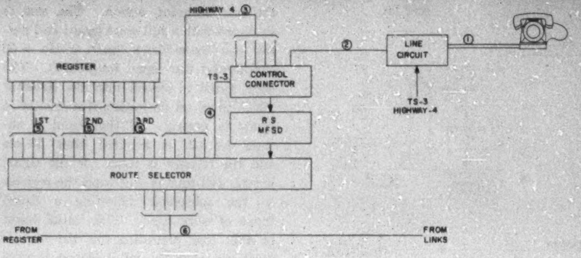

HIGHWAY 4 ③
REGISTER
1ST ⑤
2ND ⑤
3RD ⑤
TS-3
CONTROL CONNECTOR
④
R S MFSD
②
LINE CIRCUIT
①
TS-3 HIGHWAY-4
ROUTE SELECTOR
FROM REGISTER
⑥
FROM LINKS

Fig. 8. Primary and secondary busy test by route selector

1—Audio path
2—Multiplex path
3—Highway-transfer busses
4—Time-slot transfer busses
5—Digit-transfer busses
6—Busy-test busses

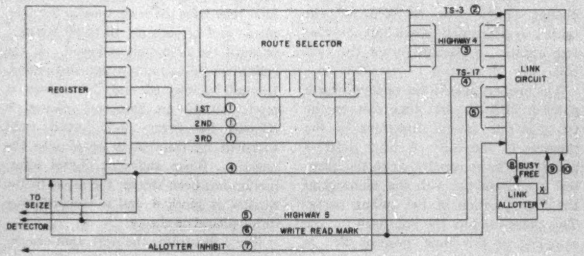

REGISTER
ROUTE SELECTOR
TS-3 ②
HIGHWAY 4 ③
TS-17 ④
LINK CIRCUIT
⑤
1ST ①
2ND ①
3RD ①
④
⑧ BUSY FREE
⑨ ⑩
LINK ALLOTTER
X
Y
TO SEIZE DETECTOR
⑤ HIGHWAY 5
⑥ WRITE READ MARK
ALLOTTER INHIBIT ⑦

Fig. 9. Transfer of calling and terminating identities from register and route selector to a link

1—Digit-transfer busses
2—Terminating time-slot transfer bus
3—Terminating highway-transfer busses
4—Calling time-slot transfer bus
5—Calling highway-transfer bus
6—Write-read mark
7—Allotter-inhibit mark
8—Link busy-free mark
9, 10—Link-selection mark

starts the trunk hunting, or selection, of the required line.

At this time, the information present on the ten busy-test busses from the registers and links is utilized in selecting a free trunk or line (Fig. 8). If a trunk is required, the busy-test busses are scanned along with time slots assigned to this trunk group. Any trunk which is busy will inhibit the appearance of a pulse within the RS and will allow the scanning action to continue. As soon as an idle trunk in the group is scanned, a pulse appears which stops the scanning action, and this pulse may be assigned to the call. A secondary busy test is performed at this time to ensure that no call is being terminated or originated externally at this time. This reduces the unguarded time between exchanges to approximately 160 μsec, or the write time between the route selector and link. In the selection of a line circuit for a subscriber, the time slot of the desired party is gated out and compared with the busy test associated with the highway to which this party is connected to determine if this party is free. If this line is found idle, the highway and time slot of this terminating party is applied to another CC identical to that previously mentioned and a seize and release detector is allowed to sample the desired subscribers lines for a period of approximately 75 ms. If either a seize signal or release signal is present, a mark is sent to the control section of the RS and a busy condition is established for the line. In place of the identity of the required party being read out, an artificial time slot and highway is transferred to the link which will allow a busy tone to be returned to the subscriber.

Assume that the line is found idle, in which case, after a prescribed sample period, the line is indicated free and the read-out action occurs. Fig. 9 indicates the elements which enter into this action. Since the identity of the calling party is stored in the register at this time, this information must be read out of the register into the link. The same information busses which were used to transfer the identity between the SD and the register are used between the register and the link. To prevent any information which may be present on these busses between the SD and register from entering a link at this time, an allotter-inhibit signal is sent from the RS erasing any information present. A guard time of at least 160 μsec follows this inhibit signal prior to the read-write mark from the RS to both the register and to all links. The LA, however, has premarked one particular link so that the read mark will only be effective in this link. The control section of the LA also senses this write-read mark and starts a timing cycle, after which a sample pulse is sent to determine if the link has read the information presented. Here again, failure to read the information is assumed as a fault in the link, and the LA steps to the next free link and attempts to engage it. Failure here also triggers a major alarm which is indicated by the control section of the LA.

Getting back to the action in the

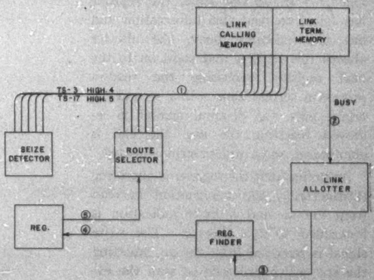

LINK CALLING MEMORY
LINK TERM. MEMORY
TS-3 HIGH. 4
TS-17 HIGH. 5
①
BUSY
②
SEIZE DETECTOR
ROUTE SELECTOR
LINK ALLOTTER
REG.
⑤
④
REG. FINDER
③

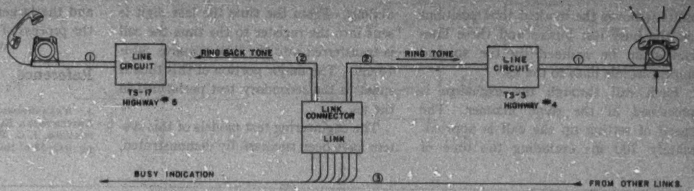
Fig. 10. Release of common equipment

1—Busy-test busses
2—Link-busy mark
3—Call-complete mark
4—Demand-answer mark
5—Digital information

Vol.
Pt. 1
195

Fig. 11. Ringing of terminating party

1—Audio paths
2—Multiplex paths
3—Busy-test busses


①
LINE CIRCUIT
TS-17
HIGHWAY # 5
RING BACK TONE
②
②
RING TONE
LINE CIRCUIT
TS-3
HIGHWAY # 4
①
LINK CONNECTOR
LINK
BUSY INDICATION
③
FROM OTHER LINKS.

RS, the write-read mark in the register will allow the identity of the calling party to be transferred from the memory of the register to a similar memory in the link. Simultaneously, on a separate set of information busses, the identity of the terminating party is written from the RS to a second memory unit in the link. The terminating party may be a line, trunk, or busy-tone equipment in the exchange. The storage of both identities is the criterion for indicating a busy link. This is forwarded to the LA on the link-free bus which starts the LA searching for another free link, and resets the answer condition of the RF (Fig. 10). The removal of the answer condition on the register resets the register and removes the digital information from the RS. Removal of the digital information resets the RS erasing the inhibit allotter and write-read mark to the links and registers. At this time the only equipment remaining with the original call is the link. The link at this time is marking two time positions on appropriate busy-test busses corresponding to the calling and terminating party identities and prevents sampling of these by the SD (Fig. 11). Since the position of the two time positions does not coincide in most cases, the multiplex signals are converted to audio in the link circuit. The conversion to audio between two different multiplex pulses will also be used in the operator's connection and detection of supervisory signals by the release, recall, and voice monitoring equipment.

When the two time slots are transferred into the link, a splitting feature prevents the audio path from being completed while a ring signal is sent to the calling party and ring-back tone to the subscriber (Fig. 11). This splitting feature is also utilized by the operator to establish recalls by releasing one or the other party without breaking down the complete connection. A ring-trip scanner samples the audio portion from each link, and is looking for a seize signal from the terminating telephone. When the called party lifts the handset from the telephone, seize signal is sent as though a call was being originated. This signal is detected by the ring-trip scanner which triggers in the link a burst of 600 cps tripping the seize tone in the terminating telephone. At the end of this burst of tone, the splitting feature in the link connects the audio path and conversation can take place between the two subscribers.

Fig. 12 indicates those equipments which are engaged in what you might call a passive roll in the exchange. The release, recall, and voice monitor is periodically gated into the conversation by the scanning action, checking to determine that conversation is in progress and that neither a recall or release signal exists. When either party terminates the call by replacing the receiver, a release signal is sent from the telephone and is detected by the release detector when it samples this link, causing a temporary storage element to be set in the link. This gates out a regenerated release signal in both directions for a period of from 2 to 4 seconds. At the end of this period, the link is reset, erasing the stored identities in the memory sections of the link and producing a free condition. Resetting of the memory

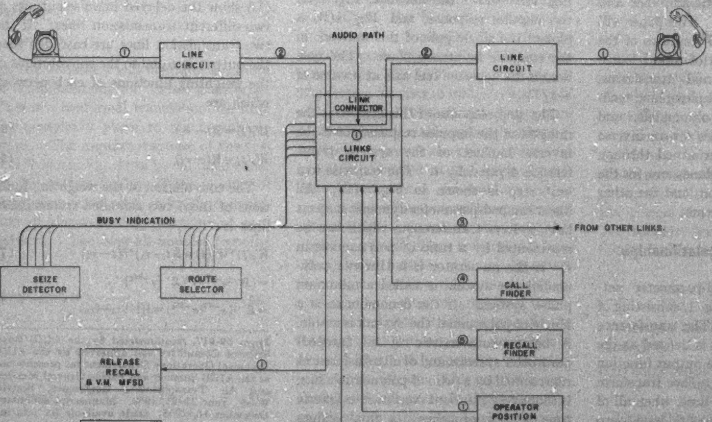

①
LINE CIRCUIT
②
AUDIO PATH
②
LINE CIRCUIT
①
LINK CONNECTOR
①
LINKS CIRCUIT
BUSY INDICATION
③
FROM OTHER LINKS.
SEIZE DETECTOR
ROUTE SELECTOR
④
CALL FINDER
⑤
RECALL FINDER
RELEASE RECALL & V.M. MFSD
①
①
OPERATOR POSITION
SCANNER GENERATOR
⑥
⑥
FROM OTHER LINKS

Fig. 12. Operating circuits during conversation period

1—Audio paths
2—Multiplex paths
3—Busy-test busses
4—Callfinder mark
5—Recall-finder mark
6—Sampling pulses

units removes the marked time positions on the busy-test busses and these lines can again be scanned by the SD and appear as idle lines to the RS.

Each call through the exchange is processed in the same manner. The speed of setting up the call is approximately 700 ms excluding the time of dialing. From the time the last digit is sent into the register to the time the call is transferred into the link requires 80± 10 ms. The major portion of this time is spent in the secondary test performed in the RS.

The engineering test models of this system have been successfully demonstrated, and the larger service test models are in the process of being tested.

Reference

1. TRANSMISSION ASPECTS OF AN ELECTRONIC SWITCHBOARD EMPLOYING TIME-DIVISION MULTIPLEXING, J. C. Perkins. *AIEE Transactions*, see pp. 949–53 of this issue.

Mixed Transcendental and Polynomial Transforms

OTTO J. M. SMITH
MEMBER AIEE

THERE ARE a variety of engineering applications in which the dynamics of the equipment being used or designed are expressible by a Laplace transform which contains both polynomial functions of s and exponential functions. In networks and in feedback control systems, transcendental terms in the Laplace transform arise due to diffusion in a transistor, dead time or time of flow of materials in heat exchangers and flow reactors, catalyst flow time in a catalytic cracker, time of transmission of an electromagnetic pulse in radar and telemetering applications, time of transmission of an ultrasonic pulse in sonar, materials transport time in thickness controls for rolling mills, the sampling time in sampled data systems and in digital computer controls, and the carrier period in magnetic-amplifier and thyratron-type controls. This paper will present some of the basic features of the mathematical treatment of mixed transcendental and polynomial transforms. Root locus methods, complex-plane techniques for factoring polynomials, and evaluation of the residues for an inverse transform can all be performed through the use of two complex planes, one for the polynomial representation, and the other for the transcendental terms.

Lumped-Parameter Relationships

The simplest lumped-parameter network is that shown in Fig. 1, consisting of an RC low-pass filter. The transference or the system transform is defined as the Laplace transform of the output function of time divided by the Laplace transform of the input function of time, when all of the energy storage elements have zero energy stored in them at time $t=0$. With this definition, the system transform is identical to the Laplace transform of the impulse response of the network or the Laplace transform of the weighting function of the network. For the circuit in Fig. 1, the weighting function is

$$W(t)=\frac{1}{T}\epsilon^{-t/T} \tag{1}$$

The Laplace transform of this is

$$G(s)=\mathcal{L}W(t)=\frac{1}{1+sT} \tag{2}$$

The sinusoidal response of this network is the value of G when $s=j\omega$. The roots of the transference $G(s)$ for complex values of s are called the zeros of the transference. The roots of the reciprocal of $G(s)$, or the roots of the denominator of $G(s)$, are called the poles of this function. Fig. 1(A) shows the network, Fig. 1(C) the impulse response, and Fig. 1(B) a plot of the single pole of this network in the complex frequency plane. The pole lies on the negative real axis at a value of $-1/T$.

The step response of this system is the integral of the impulse response, or is the inverse Laplace of the system transference divided by s. The response to a unit step is shown in Fig. 1(D). All linear lumped-parameter dynamic systems have system transferences which can be represented by a ratio of polynomials in s. If the numerator is a Hurwitz polynomial the system is called a minimum phase system. If the denominator is a Hurwitz polynomial the system is stable. A basic characteristic of all lumped-parameter systems and of all transferences represented by a ratio of polynomials in s, is that every transient continues to infinite time and approaches its final values asymptotically. Only if the initial stored energies in a system are equal to the final values required by the input transient to the system is the final value realized instantly, and there will be no transient.

Distributed Parameter Components

The characteristics of nonattenuating distortionless transmission lines are shown in Fig. 2. A transmission line with time delay τ_1 will deliver a voltage $f_1(t)$, which is equal to the input of the transmission line $f_0(t)$ delayed in time by the amount τ_1:

$$f_1(t)=f_0(t-\tau_1) \tag{3}$$

The Laplace transform of the output of a transmission line is equal to

$$\mathcal{L}f_1(t)=\epsilon^{-s\tau_1}\mathcal{L}f_0(t) \tag{4}$$

When the input to a transmission line is a unit impulse whose Laplace transform is unity, then the transform of the output of the transmission line is

$$\epsilon^{-s\tau_1}$$

This therefore, is the transference of a nonattenuating distortionless transmission line or, in fact, the transference of any pure dead time. Figs. 2(C) and 2(D) show the delayed impulse outputs of two different transmission lines. If the two transmission lines are cascaded then the output is equal to the convolutions of the weighting functions of each network which are

$$W_1(t)=\delta(t-\tau_1) \tag{5}$$

$$W_2(t)=\delta(t-\tau_2) \tag{6}$$

The convolution of the weighting functions of these two cascaded transmission lines is

$$\begin{aligned}W_1(t)^*W_2(t)&=\delta(t-\tau_1)^*\delta(t-\tau_2) \\ &=(\mathcal{L}^{-1}\epsilon^{-s\tau_1})^*(\mathcal{L}^{-1}\epsilon^{-s\tau_2}) \\ &=\mathcal{L}^{-1}(\epsilon^{-s\tau_1}\epsilon^{-s\tau_2})=\delta(t-\tau_1-\tau_2)\end{aligned} \tag{7}$$

Paper 59-917, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted December 16, 1958; made available for printing May 12, 1959.

OTTO J. M. SMITH is with the University of California, Berkeley, Calif.

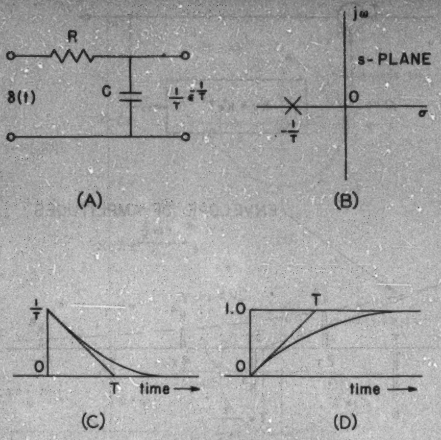


Fig. 1. Lumped-parameter system

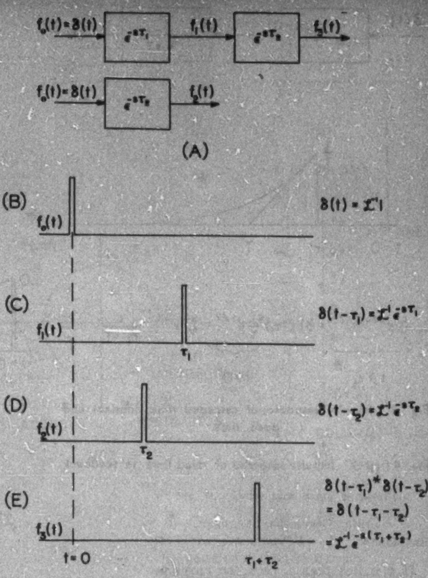

$f_0(t)=\delta(t)$
$\epsilon^{-s\tau_1}$
$f_1(t)$
$\epsilon^{-s\tau_2}$
$f_3(t)$
$f_0(t)=\delta(t)$
$\epsilon^{-s\tau_2}$
$f_2(t)$
(A)
(B)
$f_0(t)$
$\delta(t)=\mathcal{L}^{-1}1$
(C)
$f_1(t)$
τ_1
$\delta(t-\tau_1)=\mathcal{L}^{-1}\epsilon^{-s\tau_1}$
(D)
$f_2(t)$
τ_2
$\delta(t-\tau_2)=\mathcal{L}^{-1}\epsilon^{-s\tau_2}$
(E)
$f_3(t)$
t = 0
$\tau_1+\tau_2$
$\delta(t-\tau_1)^*\delta(t-\tau_2)$
$=\delta(t-\tau_1-\tau_2)$
$=\mathcal{L}^{-1}\epsilon^{-s(\tau_1+\tau_2)}$

Fig. 2 (right). Impulse responses of distortionless transmission lines, pure dead time

As shown in Fig. 2(E), this is a single impulse delayed by the sum of the delay times of the two lines.

Mixed Transcendental and Polynomial Transform

A mixed transcendental and polynomial transform is obtained by cascading a transmission line and a time constant. The Laplace transforms of the individual blocks are shown in Fig. 3. The substitution of the variable $-y$ defined by

$$-y_1=\epsilon^{-s\tau_1} \tag{8}$$

for the variable s will facilitate a variety of manipulations of these transforms. This is a conformal transformation from the complex s plane to the complex y plane. The impulse response of the system shown in Fig. 3 can be obtained either by the inverse Laplace of the product of the system transforms, or by the convolution of the weighting functions of each network. The over-all impulse response is

$$W(t)=\mathcal{L}^{-1}\epsilon^{-s\tau_1}\frac{1}{1+sT} \tag{9A}$$

$$=\delta(t-\tau_1)^*\frac{1}{T}\epsilon^{-t/T},\quad t>0 \tag{9B}$$

$$=\frac{1}{T}\epsilon^{-(t-\tau_1)/T},\qquad t>\tau_1 \tag{9C}$$

Notice that this function is zero up to the time τ_1 and then is a decaying exponential after that time. The waveform of the decaying exponential in Fig. 3 is the same as that shown in Fig. 1.

Infinite Series of Transcendental Terms

The impulse response of a system of transmission lines connected in cascade and in parallel can be only a finite number of impulses. But if the transmission line is connected in feedback, then a train of impulses extending to infinite time can be generated. This is illustrated in Fig. 4. The system consists of unity forward gain with a feedback branch including a pure dead time and the attenuation constant k. Instinctively, one knows that an impulse in will be delivered at the output immediately and that a delayed impulse of magnitude k will appear with negative sign after the time τ, due to the storage of this impulse in the feedback transmission line. The delayed impulse at time τ re-enters the transmission line, this time with a negative sign and, after an additional time τ or total time of 2τ it will pass out of the feedback path with absolute magnitude of k^2 and with a positive polarity with reference to the output. This process is continued and will generate the sequence of impulses shown in Fig. 4. The impulse response of the system can be obtained analytically by writing the inverse Laplace of the system transform which is the forward transference divided by one plus the feedback transference.

$$W(t)=\mathcal{L}^{-1}\frac{1}{1+k\epsilon^{-s\tau}}=\mathcal{L}^{-1}G_i(s) \tag{10}$$

Making the substitution of the variable y defined by equation 8, the impulse response of the system in Fig. 4 would be

$$W(t)=\mathcal{L}^{-1}\frac{1}{1-ky} \tag{11}$$

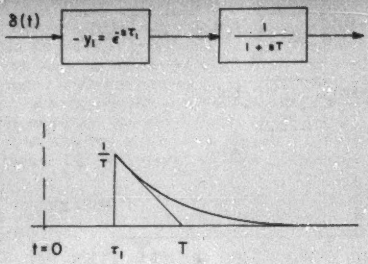
To evaluate the inverse Laplace of a ratio of polynomials in s, one uses the Heaviside partial fractions expansion to obtain a summation of low-order terms, each one of which has a simple inverse transform. In an analogous manner, to obtain the inverse Laplace of a function of y, one must obtain a summation of positive powers of y in some sort of series. The simplest series expansion for equation 11 is that for a reciprocal. It is

$$\frac{1}{1-ky}=1+ky+k^2y^2+k^3y^3+k^4y^4+\ldots \tag{12}$$

Therefore the weighting function of the network is

$$W(t)=\delta(t)-k\delta(t-\tau)+k^2\delta(t-2\tau)-k^3\delta(t-3\tau)+\ldots \tag{13}$$

This is the equation for the alternating chain of impulses shown in Fig. 4. The envelope of this chain is given by



$$\underbrace{\frac{1}{T}\epsilon^{-(t-\tau_1)/T}}_{t>\tau_1} = \delta(t-\tau_1) * \underbrace{\frac{1}{T}\epsilon^{-t/T}}_{t>0} = \mathcal{L}^{-1}\epsilon^{-s\tau_1}\cdot\frac{1}{1+sT}$$

Fig. 3. Impulse response of cascaded time constant and dead time

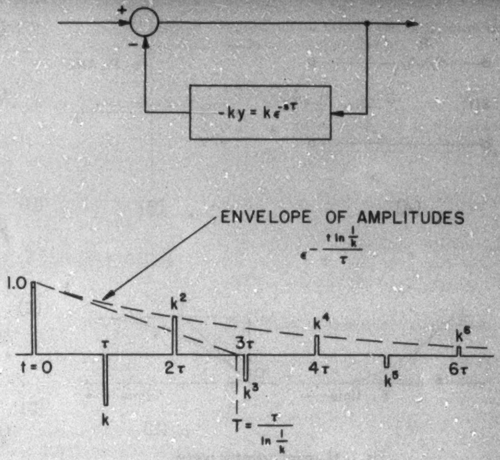

+
−
$-ky = k\epsilon^{-s\tau}$
ENVELOPE OF AMPLITUDES
$\epsilon^{-\frac{t\ln\frac{1}{k}}{\tau}}$
1.0
k^2
k^4
k^6
τ
$t=0$
2τ
3τ
4τ
6τ
k
k^3
k^5
$T = \frac{\tau}{\ln\frac{1}{k}}$

Fig. 4 (right). Impulse response of dead time in feedback

$$\mathcal{L}^{-1}\frac{1}{1-ky} = \mathcal{L}^{-1}(1+ky+k^2y^2+k^3y^3+k^4y^4+\cdots)$$

$$= \delta(t) - k\delta(t-\tau) + k^2\delta(t-2\tau) - k^3\delta(t-3\tau) + \cdots$$

$$\text{Envelope} = \epsilon^{-\frac{t\ln\frac{1}{k}}{\tau}} = k^{t/\tau} \tag{14}$$

If k is less than 1 then the envelope diminishes in amplitude. The s plane has complex pole pairs in the left-half plane. A loop gain greater than 1, however, will cause k to be greater than 1 and the envelope will continuously increase in magnitude in an unstable oscillatory manner. This corresponds to right-side s-plane poles in complex pairs.

If the feedback is positive then k is negative and a substitution in equation 13 will show that the signs of all of the impulses will be positive. The envelope will die down in a monotonic manner if k is less than 1. There is one pole on the negative real s-plane axis, but other poles as well. The envelope will increase in an unstable manner if k is greater than 1.

Fig. 5 shows the representation of the impulse response of dead time in feedback in three different complex-frequency planes. The first is the very familiar s plane. An examination of equation 10 will show that the poles of this function are given at the specific values of s of

$$s = \alpha + j\beta \tag{15}$$

when

$$\epsilon^{-s\tau} = (\epsilon^{-\alpha\tau})(\epsilon^{-j\beta\tau}) = -\frac{1}{k} \tag{16}$$

These poles occur where

$$\alpha\tau = -\ln\frac{1}{k}$$

$$\beta\tau = (2n-1)\pi \tag{17}$$

There is a column containing an infinite number of poles in the s plane, one for each value of n. These are spaced at intervals of $\beta\tau = 2\pi$. The functions plotted in Fig. 4 and Fig. 5 are for the particular case when $k = 0.707$.

Because of this infinite column of poles, the use of root locus methods in the s-plane for designing control systems is difficult. The value of a complex function is given by the distances to all of the zeros and poles,[1] for the usual rational functions.

The distance product is defined as the function of a complex variable s calculated by taking the product of all of the distances from the zeros to the value of the variable s in the s plane, divided by the product of all of the distances from the poles to the same value of the variable s in the s plane. Or, stated another way, the distance product is the continued product of all of the distances to s from every singularity, counting the distances from zeros with positive unity exponents (no change), and counting the distances from the poles with negative unity exponents (reciprocal distances). The distance product cannot be evaluated graphically when there are an infinite number of singularities. For this reason a change of variable into the y or the z plane is desirable. Fig. 5(B) shows a plot of the function $G(z)$ where the variable z is defined as

$$z = \epsilon^{+s\tau} = -\frac{1}{y} \tag{18}$$

The function must be the same, independent of changes of variable.

$$G_4(s) = G_4(y) = G_4(z) \tag{19}$$

In the z plane,

$$G_4(z) = \frac{1}{1+k\epsilon^{-s\tau}} = \frac{z}{z+k} \tag{20A}$$

Using this variable, dead time in feedback has a zero at the origin of the z plane and a pole at $-k$. The z plane is related to the s plane in the following way: the origin of the s plane maps into the point 1, 0 in the z plane. Patterns in the near vicinity of the s plane origin map unchanged into patterns in the near vicinity of (1,0) in the z plane. The $j\omega$ axis of the s plane maps into the unit circle in the z plane. The entire left-half s plane maps into the area enclosed by the unit circle in the z plane. There are an infinite number of superimposed Riemann leaves in the z plane because each horizontal strip of width 2π in the s plane maps into the z plane to cover the area of the entire plane completely.

According to equation 20(A) the function $G_4(z)$ for any point z is equal to the distance from the zero at the origin to that point, divided by the distance from

Vol.
Pt. 1
1959

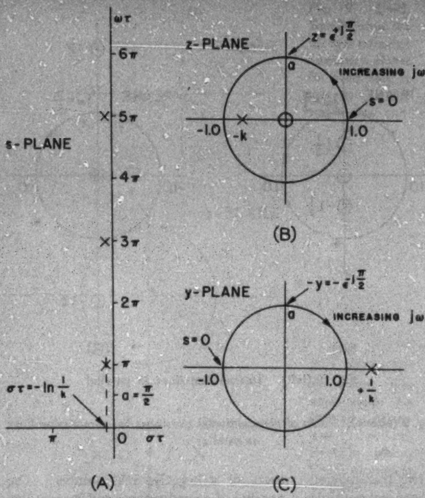


Fig. 5. **Laplace transform of impulse response of dead time in feedback**

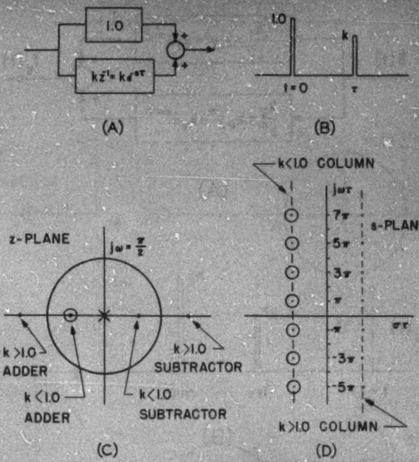


Fig. 6. **Unity plus delay in parallel**

A—Circuit used in Posicast control
B—Impulse response
C—Location of zero in z plane
D—Location of column of zeros in s plane

the pole at $-k$ to the point z. Therefore, the conventional procedure of evaluating functions of a complex variable geometrically in a complex plane by evaluating the distance product works satisfactorily in the z plane for transcendental functions of s representable by a ratio of polynomials in z.

Fig. 5(C) shows the Laplace transform of the impulse response of dead time in feedback plotted in the y plane, where y is defined by equation 8. The origin of the s plane plots into the $(-1,0)$ point in the y plane. In the vicinity of this point the s plane and the y plane are identical. The $j\omega$-axis of the s plane plots into the unit circle in the y plane and the point $j\pi/2$ marked a in the s plane plots into the point $j1$ in the y plane. The entire right half of the s plane is enclosed by the unit circle in the y plane. All singularities outside of the unit circle correspond to stable poles or minimum-phase zeros in the s plane. The y plane has many Riemann leaves, each horizontal strip of vertical width 2π in the s plane plots into a complete coverage of the y plane. One advantage of using the y plane is that the infinite column of poles in the s plane becomes a single pole in the y plane. The function is given by

$$G_4(y)=\frac{1}{1+k\epsilon^{-s\tau}}=\frac{1}{1-ky}=\left(\frac{-1}{k}\right)\frac{1}{\left(y-\frac{1}{k}\right)} \qquad (20B)$$

This has a single pole located at $y=1/k$.

The function $G_4(y)$ can be evaluated geometrically in the y plane by first factoring out the constant $-1/k$ as shown in equation 20(B). The remaining factors of y have a distance product equal to the product of the distances from all of the zeros to the point y divided by the product of the distances from all the poles to the point y. Again this function of a complex variable can be evaluated geometrically.

When there are no zeros, use unity for the "distance to the zero." It is apparent that, to perform this calculation, one may choose whichever plane results in the simplest computation. The s plane is ruled out because of the infinite column of poles. In this particular example the y plane is most convenient because it has a single pole. However, the z plane is not much more difficult to use, because it has only one pole and one zero.

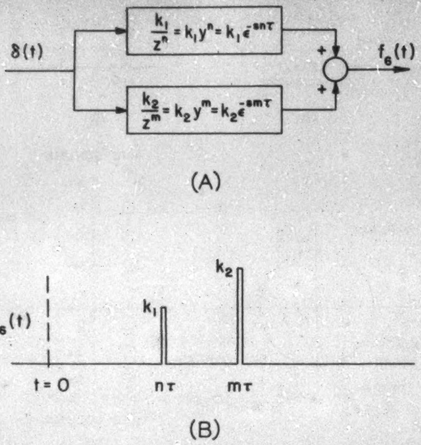
Fig. 5(A) has an infinite column of minimum-phase zeros which lie at negative real infinity at the far left of the s plane. If the output of the circuit in Fig. 4 had been taken from the output of the transmission line, then the first pulse of the pulse train in Fig. 4 would have occurred at time τ. This would have been a nonminimum phase system, and the infinite column of zeros for Fig. 5(A) would have been at the far right edge of the s plane. The z plane would not have had a zero at the origin, and the y plane would have had a zero at the origin.

Polynomial of Transcendental Terms

Two transmission lines in parallel, or a transmission line in parallel with a constant transference, will have a finite number of impulses in the impulse response.

An infinite train of pulses in the time domain must be represented by poles in the frequency domain. If the train is composed of impulses there will be an infinite number of s-plane poles, or at least one z-plane or y-plane pole. This is analogous to the time domain responses shown in Fig. 1 which approach a final value asymptotically at infinite time, and therefore must be represented by a pole, in this case, in the s plane.

If the time domain response stops in finite time, then the Laplace transform does not have any poles in the finite s plane. It also does not have any poles in the finite y plane, nor can a pole lie within the unit circle which would represent an unstable system. The transform can have poles in the finite z plane however, because poles at the origin of the z plane are necessary to balance all of the zeros which appear in the finite z plane.


δ(t)
$\frac{k_1}{z^n} = k_1 y^n = k_1 \epsilon^{-sn\tau}$
$\frac{k_2}{z^m} = k_2 y^m = k_2 \epsilon^{-sm\tau}$
+
+
$f_6(t)$
(A)
$f_6(t)$
k_1
k_2
t = 0
nτ
mτ
(B)

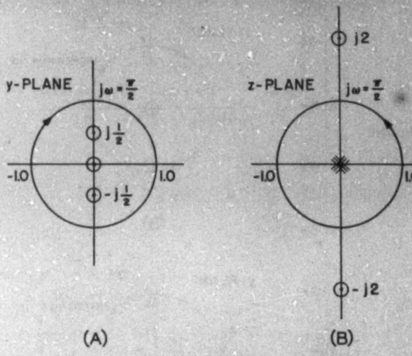


Fig. 7 (left). Transmission lines in parallel

Fig. 8 (above). Roots of transcendental transforms for transmission lines in parallel

Fig. 6 shows a transmission line added to unity. In this case the gain coefficient of the transmission line k has been selected as less than unity for purposes of illustration. Fig. 6(B) shows the impulse response consisting of two impulses only. The Laplace transform of the impulse response of this system is given by

$$G_5 = \mathfrak{L}[\delta(t) + k\delta(t-\tau)]$$
$$= 1 + kz^{-1} = \frac{z+k}{z} \qquad (21)$$

This function has one pole and one zero in the z plane as shown in Fig. 6(C). The pole is at the origin. The zero lies within the unit circle if k is less than 1. This corresponds to a minimum-phase system. If k is greater than 1, the second impulse is larger than the first impulse, and the zero of the transference lies outside the unit circle. This corresponds to a non-minimum phase system. The z-plane zero is plotted in the s plane in Fig. 6(D). It becomes an infinite column of zeros at $(\alpha\tau = -ln\ 1/k)$, with vertical co-ordinates of the zeros being the values of $\beta\tau$ given in equation 17. If $k > 1$ then the real co-ordinate of the column becomes positive and the column of zeros is in the right-half s plane.

G_5 as given in equation 21 can be evaluated for any value of s by finding the corresponding value of z, and then computing the distance product which, in this case, is the distance from the zero to the point z in the z plane divided by the distance from the origin to the point z in the z plane.

A more general case is that shown in Fig. 7, where the system consists of two transmission lines in parallel. The delay time is different in the two transmission lines. A minimum common submultiple of the two delay times is chosen. This value is called τ. The delay time of the first line is $n\tau$ and the delay time of the second line is $m\tau$. Their Laplace transforms are

$$G_n = k_1 \epsilon^{-sn\tau}$$
$$G_m = k_2 \epsilon^{-sm\tau} \qquad (22)$$

The impulse response of these two transmission lines in parallel is given in Fig. 7(B). It is

$$f_6(t) = k_1\delta(t-n\tau) + k_2\delta(t-m\tau)$$
$$= \mathfrak{L}^{-1}(k_1\epsilon^{-n\tau s} + k_2\epsilon^{-m\tau s}) \qquad (23)$$

Fig. 8(A) shows the y-plane plot for the system transference of the two transmission lines in Fig. 7, derived from

$$\mathfrak{L}f_6(t) = k_1(-y)^n + k_2(-y)^m$$
$$\mathfrak{L}f_6(t) = k_1(-y)^n\left[1 + \frac{k_2}{k_1}(-y)^{m-n}\right]$$
$$= k_2(-y)^n\left[\frac{k_1}{k_2} + (-y)^{m-n}\right] \qquad (24)$$

The transference has an nth order zero at the origin of the y plane and a large number of other zeros at the roots of

$$y_0 = -\left(-\frac{k_1}{k_2}\right)^{\frac{1}{m-n}} \qquad (25)$$

The particular case plotted in Fig. 8(A) is for $n=1$, $m=3$, $k_1=1$, and $k_2=4$. For this particular case, the complex zeros are nonminimum phase because $k_2 > k_1$. There are two complex zeros because the difference between m and n is 2. In general the complex zeros will be equiangularly located on a circle corresponding to the roots of a negative real number. The number of complex zeros will be equal to the difference between m and n. There will be, in addition, a high-order zero at the origin of the y plane with an order corresponding to the smaller of the two numbers n or m. The total number of zeros in the y plane is equal to the larger of the two numbers n or m. In a similar manner, the z-plane pattern can be plotted from

$$\mathfrak{L}f_6(t) = \left(\frac{k_1}{z^m}\frac{k_2}{k_1} + z^{m-n}\right) = \frac{k_2}{z^m}\left(1 + \frac{k_1}{k_2}z^{m-n}\right) \qquad (26)$$

This has an m-order pole at the origin of the z plane and it has complex zeros located at the roots of

$$z_0 = \left(-\frac{k_2}{k_1}\right)^{\frac{1}{m-n}} \qquad (27)$$

These complex zeros will be equiangularly spaced on a circle centered at the origin of the z plane. The transference of the circuit in Fig. 7(A) can be evaluated for any complex frequency by determining the corresponding value of y or z for this complex frequency, and determining the distance product in the y or the z plane. It is important that the gain coefficient k_2 or k_1 respectively be multipled by the distance product to obtain the value of the transference.

Mixed Lumped Parameters and Dead Time

The systems shown in Figs. 4 through 8 inclusive are all composed of pure dead time components. Their transferences

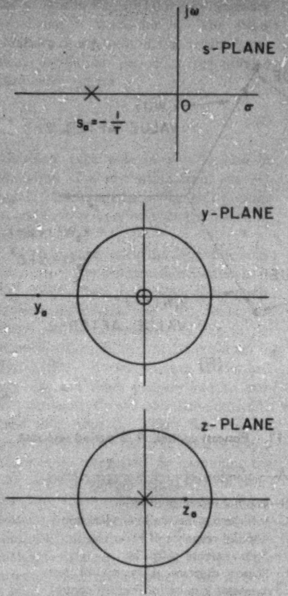


Fig. 9 (left). Simultaneous use of two complex frequency planes for dead time in cascade with single time constant

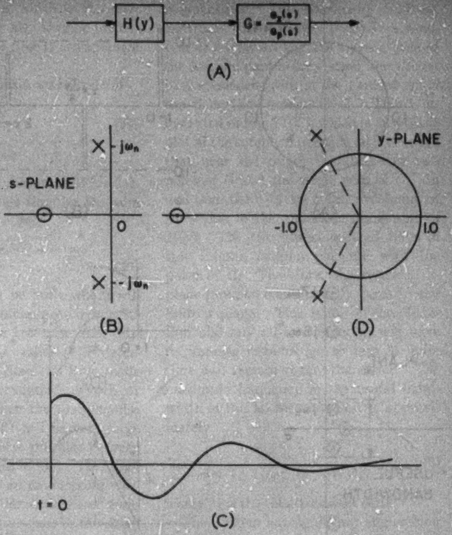


Fig. 10 (right). A—Delay-line compensator for lumped-parameter system; B—G in s plane; C—G plotted in y plane; D—Impulse response of G alone

are polynomials in y or z. The only mixed system which has been discussed so far is that shown in Fig. 3. Its transference is the product of a function of y and a function of s. It is interesting to examine how this function may be expressed graphically. The transference is

$$\frac{\epsilon^{-s\tau_1}}{1+sT} = \frac{-1}{T}\, y\, \frac{1}{s+\frac{1}{T}}$$

$$= \frac{1}{T}\,\frac{1}{z}\,\frac{1}{s+\frac{1}{T}} \tag{28}$$

The equation was arranged so that only positive powers of the variable appeared. To represent this system with a finite number of poles and zeros requires two different complex-frequency planes. One of these must be the s plane and the other can be either the y plane or the z plane. In the s plane there is a single pole at $-1/T$. In the z plane there is a pole at the origin. Fig. 9 shows these two planes and the locations of their singularities. In addition, the location of the pole at s_a is plotted in the z plane at the point

$$z_a = \epsilon^{+s_a\tau} \tag{29}$$

In Fig. 9 the y plane is shown. It has a single zero at the origin. In addition, a point is shown on the y plane corresponding to the location of the pole in the s plane. This point y_a is given by

$$y_a = -\epsilon^{-s_a\tau} \tag{29A}$$

It may seem trivial to evaluate the impulse response of this system from the complex frequency planes shown in Fig. 9. However, for more complex systems, this is an efficient and practical way to obtain the impulse response. The impulse response is the convolution of the chain of impulses generated by the function of y and the summation of exponentials generated by the function of s. In this trivial case, the s plane yields a single term which is

$$\frac{1}{T}\epsilon^{-t/T}$$

Also in this particular case a single zero in y produces the function $-\delta(t-\tau_1)$. The resultant impulse response of the system is therefore

$$W(t) = \left(-\frac{1}{T}\right)\left(-\delta(t-\tau_1)^{*}\epsilon^{-t/T}\right) \tag{30}$$

The convolution of a finite sum of impulses and a continuous function is best evaluated by the application of the superposition theorem. In this particular case, there is only one impulse and the continuous function is only shifted in time as indicated in equation 9(C)

$$W(t) = \frac{1}{T}\epsilon^{-(t-\tau_1)/T} \quad t > \tau_1$$

Posicast Control

There is a form of control called a Posicast system.[1] These systems have a delay-line compensator prior to a lumped-parameter system. For a single impulse into the compensator, it delivers a chain of impulses of finite number and these excite the lumped-parameter system in such a manner that it reaches its final value in finite time and the output of the lumped-parameter system remains constant at its final value thereafter. These are called deadbeat systems by engineers. The engineering features have been described in the literature.[2,3] To get an intuitive feeling for this method of control imagine that the lumped-parameter system has a significant oscillation with small damping. Then each impulse that leaves the delay-line compensator and excites the lumped-parameter system will excite a damped oscillation. The condition for satisfactory deadbeat control is that the sum of the various shrinking phasors representing these damped oscillations should add to zero after the last impulse excitation. The algebraic sum of the areas of the several impulses that entered the lumped-parameter system should be unity to correspond with the unit impulse which enters the delay-line compensator.

Another way of looking at this type of control is to imagine that the delay-line

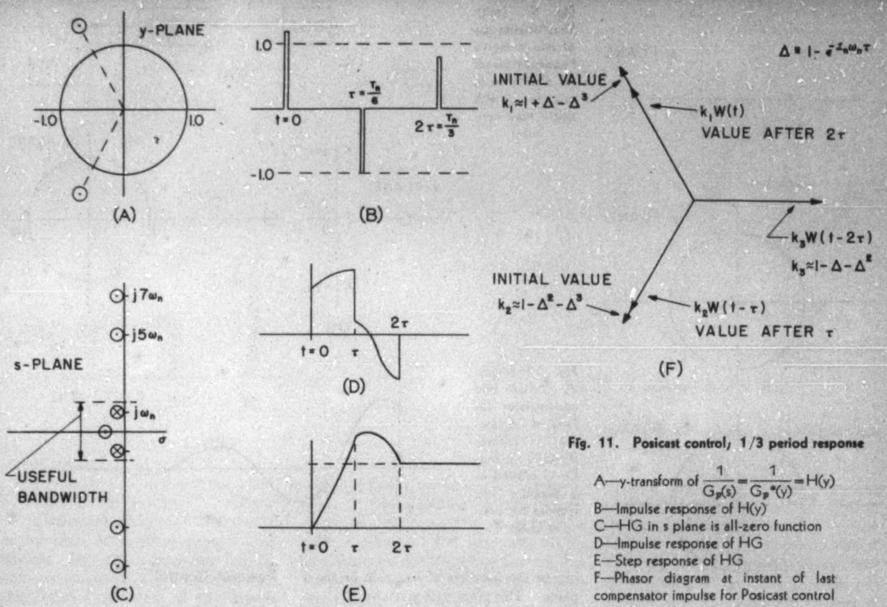


Fig. 11. Posicast control, 1/3 period response

A—y-transform of $\frac{1}{G_p(s)}=\frac{1}{G_p{}^*(y)}=H(y)$

B—Impulse response of H(y)

C—HG in s plane is all-zero function

D—Impulse response of HG

E—Step response of HG

F—Phasor diagram at instant of last compensator impulse for Posicast control

compensator generates a polynomial in y. This polynomial in y has zeros in the complex-frequency plane. The lumped-parameter system however has poles which represent its resonant frequencies and time constants. The correct deadbeat design is one in which, for each pole in the lumped-parameter system there is a coincident zero generated by the delay-line compensator.

Fig. 10 shows an example of this type of system. Fig. 10(A) is the block diagram of the compensator and the oscillatory member. Fig. 10(B) shows the transfer function of the oscillatory member in the s plane. For purposes of example, this is taken to be a pair of resonant poles and a single real zero. Fig. 10(C) shows the impulse response of G alone. It has an initial step because the system has only one more pole than the number of zeros.

The locations of the poles and zeros of G in the s plane can be plotted at the corresponding complex frequencies in the y plane. We will define a new function of y called the y transform of $G(s)$, which is usually written as $G^*(y)$. This new function, $G^*(y)$, has poles and zeros in the y plane at the identical complex frequencies for which there were poles and zeros of G in the s plane. These locations can be found from the plot of G in the y plane. The function $G^*(y)$ is not exactly equal to $G(s)$. If the function $G^*(y)$ is plotted in the s plane, as it can be, then it would consist of an infinite column of poles in the s plane and another infinite column of zeroes. Only the very-low-frequency pair of poles of this column coincide in the s plane with the poles of G and only one of the zeros of the column of $G_z{}^*(y)$ plotted in the s plane coincides with the zero G_z.

To find any y-transform function $F^*(y)$ given the corresponding s-transform $F(s)$, one may substitute for each term in the s transform of the form

$$\frac{s-s_1}{0-s_1}$$

a corresponding term in the y transform equal to

$$\frac{\epsilon^{-s\tau}-\epsilon^{-s_1\tau}}{\epsilon^{0}-\epsilon^{-s_1\tau}}=\frac{-y+y_1}{1+y_1}=\frac{y_1-y}{y_1+1} \tag{31}$$

The original function $G(s)$ is divided into a ratio of polynomials

$$\frac{G_z(s)}{G_p(s)}$$

The poles

$$\frac{1}{G_p(s)}$$

are plotted in the y plane as the y transform

$$\frac{1}{G_p{}^*(y)}$$

For the location of each root of $G_p{}^*(y)$ in the y plane one generates a y-plane zero. The product of all of these zeros is a function $H(y)$ given by

$$H(y)=\frac{1}{G_p{}^*(y)} \tag{32}$$

This is a polynomial in positive powers of y. It has only zeros in the y plane. There are no poles. Each zero of $H(y)$ is coincident with a pole of the function $1/G_p{}^*(y)$, which is the y transform of all of the poles of the lumped-parameter system G. Fig. 11(A) shows the necessary zeros in the y plane corresponding to the system shown in Fig. 10.

The zeros in Fig. 11(A) can be generated with a quadratic in y. The three terms of this quadratic are shown in Fig. 11(B) as a sequence of impulses. The first impulse is slightly greater than 1; the second impulse is approximately equal to 1; and the third impulse is slightly less than 1 for the particular value of τ which has been chosen for the definition of y. In this particular case, in order to plot the s-plane function in Fig. 10(B) into the y plane in Fig. 10(D) it was necessary to choose a value for τ. The value chosen was one sixth of the natural transient oscillation period of the lumped parameter system.

$$\tau = \frac{T_n}{6} = \frac{\pi}{3\omega_n} \tag{33}$$

Therefore each pole in y is located at one sixth of a revolution from the -1 point. Each impulse in Fig. 11(B) occurs at one sixth of a period after the preceding impulse.

Each impulse from the delay-line compensator excites a damped oscillation as in Fig. 10(C), which approaches zero asymptotically for infinite time. The sum of the three oscillatory components due to the three impulses in Fig. 11(B) is shown in Fig. 11(D). At the instant of the last impulse, the three phasors add to zero and consequently the oscillation ceases to exist after time 2τ. That this impulse response is a very desirable mode of operation can be recognized by examining the step response of the system with its compensator, Fig. 11(E). The output of the lumped-parameter system for a unit step into the compensator is given by the integral of the curve in Fig. 11(D). This integral rises monotonically for a time a little greater than τ until it reaches its peak and then it diminishes a small amount until the time 2τ, after which it remains constant thereafter. The reason why the step response of this system shows an overshoot is because of the zero of G in the s plane, due to the transference of the lumped-parameter system. The system response reaches its final value in 1/3 cycle.

The phasor components of the oscillation are shown in Fig. 11(F). The first impulse excites the oscillation $k_1W(t)$ which has a radian frequency ω_n and dimensionless damping ζ_n. A measure of the damping is the per-unit envelope loss in 1/6 cycle,

$$\Delta = 1 - \epsilon^{-\zeta_n\omega_n\tau} \tag{34}$$

The first impulse has a magnitude of

$$k_1 = \frac{1}{(1-\Delta)^2+\Delta} \tag{35}$$

After 2τ, this phasor has an angle of 120 degrees and has shrunk (attenuated) to

$$k_3 = (1-\Delta)^2 k_1 \tag{36}$$

The second impulse excites the oscillation $-k_2W(t-\tau)$, which has an initial magnitude of

$$k_2 = (1-\Delta)k_1 = \frac{1}{(1-\Delta)+\left(\dfrac{\Delta}{1-\Delta}\right)} \tag{37}$$

After τ units of time, which is at $t=2\tau$, this phasor will have also shrunk (attenuated) to k_3. At this time it has an angle of -120 degrees.

The last impulse has a magnitude of

$$k_3 = \frac{1}{1-\dfrac{\Delta}{(1-\Delta)^2}} \tag{38}$$

and excites the phasor $k_3W(t-2\tau)$. Immediately after this impulse the sum of the three phasors is zero.

$$k_1W(t) - k_2W(t-\tau) + k_3W(t-2\tau) = 0 \qquad t > 2\tau \tag{39}$$

Other designs can be built which will cause the system response to reach its final value in more or less time than this. The selection of the value of τ determines the response time. If τ is made very small, as for example, 1/20th or 1/30th of a cycle, then the first impulse required in Fig. 11(B) will be very large and the second impulse will also be very large in a negative direction. This may not be feasible from an engineering view point, because these large values of compensation signal require gain in the delay line compensator and in addition may saturate linear components of the system.

Complex-Zero Generator

The principle of the Posicast control described in Figs. 10 and 11 can be used to test unknown systems. The signal generator for the test delivers two successive step functions, like the integral of two successive impulses. The ratio of amplitudes and the time spacing are adjustable. The signal generator is connected to a system, and the output is observed on an oscilloscope. The amplitude and pulse timing are varied until the oscilloscope shows a deadbeat response.

The signal generator has produced an infinite column of complex zeros in the s plane, and deadbeat response means that two of these zeros are coincident with the two complex poles of the unknown system. The resonant frequency and the damping of the system poles can then be read directly from the dials of the signal generator, which are calibrated in terms of the locations of the complex zeros.

This is analogous to a Q-meter, except that the accuracy is high for low Q's, and diminishes for increasing Q.

All-Zero Functions

It can be proved that deadbeat response can be achieved with any lumped-parameter system preceded by a delay line compensator. The criterion for deadbeat control is that for every pole of the function of s representing the transference of the lumped-parameter system, there must be a coincident zero in the y transform of the delay line compensator. This is illustrated in Fig. 11(C), which is an s-plane plot of the complete function HG. Notice that near the origin there are the two poles of G and the single zero of G. In addition there is an infinite column of zeros due to the quadratic numerator of $H(y)$. The two lowest ordered zeros in this infinite column coincide with the poles of G. Therefore, the resultant s-plane function contains only zeros in the finite s plane. This is an all-zero function and any all-zero function will have its impulse response go to zero in finite time and remain at zero thereafter. The maximum frequency in the useful bandwidth of this all-zero function is approximately

$$f_{\max} = \frac{1}{4\tau} \tag{40}$$

where 2τ is the total transient time.

This system has significant engineering advantages over the usual Chebyshev or Butterworth designs for a uniform gain throughout a bandwidth. In these conventional designs, the useful frequency band is filled with poles which distort the function and produce ripples in both gain and phase. Outside of the useful frequency band the function is quite smooth and well-behaved although near zero in magnitude. In the Posicast design the useful frequency band is devoid of singularities and consequently the function has very uniform and constant gain throughout the bandwidth and it is only in the stop band or at the very high frequencies that there are ripples in the gain function, but these occur at frequencies where the gain is so low that the variations in the gain are negligible.

A generalized Posicast control system has an over-all transfer function given by

$$HG = K\prod_{i=1}^{n}\frac{y-y_i}{s-s_i} \tag{41}$$

The zeros of the lumped-parameter transference have been neglected. Each pole of the lumped-parameter transference has a corresponding zero of the y transform at the same complex frequency. All of the constants have been lumped together in the factor K. The specification that each zero of y coincide with a pole of s is given by

$$y_i = -\epsilon^{-s_i\tau} \tag{42}$$

Making this substitution, the system transference is

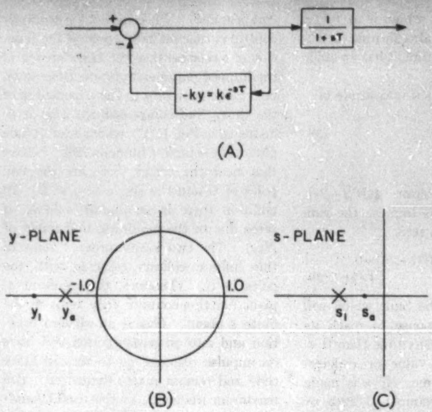


Fig. 12 (left). Simultaneous use of two complex frequency planes for dead time in feedback cascaded with a single time constant

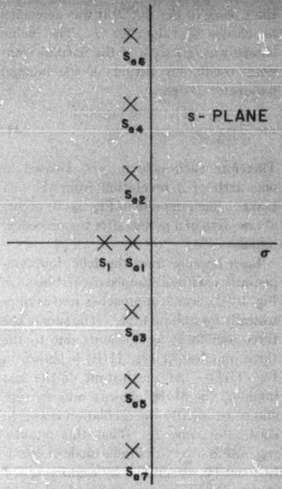


Fig. 13 (right). Poles of partial-fractions expansion of a transform in s and y

$$HG=K\prod_{i=1}^{n}\frac{\epsilon^{-s_i\tau}-\epsilon^{-s\tau}}{s-s_i} \tag{43}$$

It can be shown that the convolution of pulses, each of which lasts only for a finite time, is a resultant pulse which also lasts only for a finite time. To prove that this over-all transference is deadbeat, therefore, we need only prove that each component is deadbeat. By a component is meant a pair consisting of one s-plane pole and its corresponding y-plane zero.

The inverse Laplace of one component of this continued product is given by

$$\begin{aligned}\mathfrak{L}^{-1}C_i&=\mathfrak{L}^{-1}\left(\frac{1}{s-s_i}\right)(\epsilon^{-s_i\tau}-\epsilon^{-s\tau})\\&=\left(\mathfrak{L}^{-1}\frac{1}{s-s_i}\right)^{*}[\mathfrak{L}^{-1}(\epsilon^{-s_i\tau}-\epsilon^{-s\tau})]\end{aligned} \tag{44}$$

The inverse Laplace of the product of two transforms is equal to the convolution of the inverse Laplaces of each transform alone. The first inverse Laplace yields an exponential decay term valid only for positive time. The second inverse Laplace yields the summation of an impulse at zero time and a delayed impulse. The convolution of these two functions is the time domain contribution of this one component pair.

$$\mathfrak{L}^{-1}C_i=\underset{t>0}{\epsilon^{+s_it}}{}^{*}(\epsilon^{-s_i\tau}\delta(t)-\delta(t-\tau)) \tag{45}$$

This contribution is zero for negative time and has the value stated only for positive time. The convolution of the sum of two functions can be written as the sum of the convolutions, and this results in

$$\mathfrak{L}^{-1}C_i=\underset{t>0}{\epsilon^{+s_it}}{}^{*}\epsilon^{-s_i\tau}\delta(t)-\epsilon^{+s_it}{}^{*}\delta(t-\tau) \tag{46}$$

With regard to the first term above, the convolution of a function of time with an impulse at the origin does not change the function of time. Therefore the first convolution is only multiplication, and the function of time still has value only for time greater than zero.

The second term in equation 46, however, is the same as the action of the system shown in Fig. 3 and given by equation 9. It is the convolution of a function of time valid for positive time only with a delayed impulse. The effect of the delayed impulse is to shift the time scale so that the given function of time starts at a later time, and is zero up to this later time and has value only after time τ. Therefore, the sum of the two functions in equation 46 are

$$\mathfrak{L}^{-1}C_i=\underset{t>0}{\epsilon^{+s_i(t-\tau)}}-\underset{t>\tau}{\epsilon^{+s_i(t-\tau)}} \tag{47}$$

The first function starts at time $t=0$ and continues for all time. The second function starts at time $t=\tau$ and continues for all time thereafter. Since the two functions are identical the difference exists only for the time between zero and τ.

$$\mathfrak{L}^{-1}C_i=\epsilon^{+s_i(t-\tau)}\quad 0<t<\tau \tag{48}$$

and the difference is equal to zero for all times that lie outside of this finite span

$$\mathfrak{L}^{-1}C_i=0\quad t<0,\ t>\tau, \tag{49}$$

If there are n different terms in equation 37 then the convolution of the n resultant pulses, each one lasting for time τ, will be a function which lasts for time $n\tau$.

$$\mathfrak{L}^{-1}HG=0\quad t<0,\quad t>n\tau \tag{50}$$

This is the desirable characteristic which Posicast control achieves.

Graphical Residue Evaluation

A system consisting of a delay line compensator with feedback followed by a lumped-parameter system with feedback can have poles both in the y plane and the s plane. To evaluate the characteristics of this kind of system, it is convenient to have rapid graphical methods of determining the frequency response of the system, the root locus in case an additional feedback path is introduced around the cascaded combination and the residues at the poles, so that the impulse and step response may be written down directly from graphical evaluations, rather than from analytical evaluations.

To illustrate these techniques, we will consider the system shown in Fig. 12. This has a delay line in feedback followed by a single-time-constant lumped-parameter system. This has a transfer function given by

$$G_s=\frac{1}{(1+ky)}\frac{1}{(1+sT)}=\frac{1}{Tk\left(y+\frac{1}{k}\right)\left(s+\frac{1}{T}\right)} \tag{51}$$

To distinguish the two kinds of poles and zeros as well as the two kinds of frequency planes, we will call all poles and zeros due to lumped-parameter elements as black ones, and all poles and zeros due to distributed-parameter elements as red ones. We will use subscript numbers for the black poles and subscript letters for the red poles. A single red pole in the y plane plots into an infinite column of red-pole frequencies in the s plane. A black pole in the s plane plots into a single black-pole frequency in the y plane.

The transference of the one black pole and one red pole in equation 51 can be written

$$\frac{G_8}{K}=\frac{1}{(s-s_1)(y-y_a)} \qquad (52)$$

The inverse transform of this is a transient which extends through all time. However, because of the attenuating character of the lumped-parameter filter, the frequency components in this transient have a limited bandwidth. Thus, an acceptable method of writing the transient would be to determine the significant Fourier components and write only these. This means that the single red y-plane pole is replotted as a column of poles in the s plane and the residues are evaluated for only the significant low-frequency poles in this column.

Fig. 12(B) shows the single red y-plane pole at $-y_a$ and Fig. 12(C) shows the single black s-plane pole at $-s_1$. Fig. 13 shows the red poles due to the transmission line plotted in the s plane and marked s_{a1}, s_{a2}, s_{a3}, etc.

The first step in evaluating the inverse transform of G_8 is to expand it in partial fractions corresponding to the terms shown by the poles in Fig. 13. This partial fractions expansion is

$$\frac{G_8}{K}=\frac{R_1}{(s-s_1)}+\frac{R_{a1}}{(s-s_{a1})}+\frac{R_{a2}}{(s-s_{a2})}+\frac{R_{a3}}{(s-s_{a3})}+\dots \qquad (53)$$

This series might not converge, but if the higher frequencies are attenuated, only low-frequency poles need be considered.

The residues are the numerators in this partial-fractions expansion. The residue at the black pole is

$$R_1=\lim_{s\to s_1}\frac{G_8}{K}(s-s_1)=\frac{1}{(y_1-y_a)} \qquad (54)$$

The reciprocal residues at the red poles are

$$R_{an}{}^{-1}=\lim_{s\to s_{an}}\frac{d}{ds}(y-y_a)(s-s_1)$$

$$R_{an}{}^{-1}=\lim_{s\to s_{an}}(y-y_a)+(s-s_1)y(-\tau) \qquad (55)$$

$$R_{an}{}^{-1}=(s_{an}-s_1)y_a(-\tau)$$

In general, the residues at the red poles are

$$R_{an}=\lim_{s\to s_{an}}\frac{G_8(y-y_a)}{Ky_a(-\tau)} \qquad (56)$$

The resultant partial-fractions expansion, which is in a form directly amenable to using inverse Laplace Tables, and in which each term represents a particular frequency in the s plane, is

$$\frac{G_8}{K}=\frac{1}{(y_1-y_a)(s-s_1)}+\frac{1}{(-\tau)}\sum_{n=1}^{\infty}\frac{1}{(s-s_{an})(s_{an}-s_1)y_a} \qquad (57)$$

The residues are the coefficients of the time terms in the inverse Laplace. The magnitudes of these coefficients can be obtained directly from the geometry in Fig. 12. For example, R_1 is the residue of the single s-plane pole due to the lumped-parameter network. If this were completely a lumped-parameter system, then the residue at s_1 would be the value of the function at s_1 with the pole there removed. This is conventionally calculated as the distance product to s_1 from all the other poles and zeros in the system. But as can be seen in Fig. 13, the number of other poles in the s plane is infinite and this distance product would be difficult to evaluate in the s plane. However, equation 54 shows that this distance product can be evaluated directly in the y plane. The residue R_1 for the black pole is equal to the reciprocal of the distance in the y plane from the red pole at y_a to the complex frequency point y_1. This distance is therefore the coefficient of the term

$$\epsilon^{-t/T}$$

in the impulse response of the given system.

The residue at the red pole s_{a1} in the s plane is given by equation 55 or 56. The residue at the red pole frequency s_{a1} in the s plane is equal to the reciprocal of the distance from the black pole s_1 to the location s_{a1} where the residue is to be evaluated, times the distance in the y plane from the origin to the red pole y_a, times $(-\tau)$. This residue is the value of the function G_8/K at the frequency s_{a1}, with the red pole at s_{a1} removed.

There is a calibration constant which has to be included when the residue at a red pole is evaluated in the s plane. This calibration constant is the negative reciprocal of the time base for the y plane, $(-1/\tau)$.

The technique just illustrated can be generalized as follows: A transference which is a product of a ratio of polynomials in s times a ratio of polynomials in y can be expanded in a sum of partial fractions with all of the denominators having units of s. The residue at any given black pole is the distance product from all of the black singularities in the s plane times the distance product from all of the red singularities in the y plane, to the complex frequency of the given black pole. The residue at any given red pole frequency in the s plane is equal to minus $1/\tau$ times the distance product from all of the black singularities, in the s plane, times the distance product from all of the red singularities in the y plane, with an extra pole at the origin of the y plane.

Conclusions

It is hoped that this exposition will have demonstrated the usefulness of a variety of mathematical transformations and techniques. These introductory relationships between frequency domain and time domain representations of transforms which have mixed polynomial and transcendental terms, should be extended to form a complete unified treatise.

An important phase of this work concerns the direct evaluation of residues and root loci in the complex frequency plane, for systems involving feedback around cascaded delay compensators and lumped-parameter elements. Another important field which needs additional work is that involving diffusion processes such as occur in transistors. These transcendental terms have transforms of the form

$$\epsilon^{-\sqrt{s\tau}}$$

Here again a new frequency-domain plane is needed. I hope that these problems will be intriguing to engineers and that they will contribute significantly to our understanding and to the tools available in these fields.

References

1. Feedback Control Systems (book), Otto J. M. Smith. McGraw-Hill Book Company, Inc., New York, N. Y., 1958, p. 35 and pp. 331–44.

2. Posicast Control of Damped Oscillatory Systems, O. J. M. Smith. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 45, Sept. 1957, pp. 1249–55.

3. Analogue Study of Deadbeat Posicast Control, George Tallman, Otto J. M. Smith. *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, New York, N. Y., vol. PGAC-4, March 1958, pp. 14–21.

4. Mixed Distributed and Lumped Parameter Systems, Otto J. M. Smith. *WESCON Convention Record*, Institute of Radio Engineers, New York, N. Y., vol. 1, pt. II, 1957, pp. 122–32.

Simulation of Data-Switching Systems on a Digital Computer

F. J. GROSS
MEMBER AIEE

SINCE 1941 automatic data-message-switching systems using teletypewriter techniques have been widely used to transmit messages from each of a multiplicity of sources to destinations in accordance with address codes included in each message.[1] Existing systems have increased in size and complexity along with the business operations which they serve. Future data-message-switching systems show promise of being even larger and more complex, and of operating at much higher speeds.

The performance of such systems is measured by the average annual cost per message for specified requirements of speed-of-message delivery and peak hour capacity under specified conditions of average message length, geographical distribution of sources and destinations, and ratio of annual load to peak-hour load. Maximum performance depends not only on the sound design of the elements of the system, but also upon the skill with which the data system is adapted to the business operations of the user.

Judgment as to the ability of these systems to meet requirements of capacity and speed of delivery involves the laws of statistics and probability that become more difficult to apply as the system becomes more complex.

Simulation of the operation of a data-switching system on a large digital computer is a powerful tool for determining how well it will meet the capacity and speed of service requirements for any specified application. By cut-and-try methods of simulating the operation of various designs and system layouts, one can select the most economical one that will adequately meet the requirements.

Furthermore, if digital computers or other automatic devices are to be directly connected to the data-switching system their operating parameters can be included in the simulation to determine the performance of the proposed composite system.

Equations and graphs based on the theory of probability, such as the telephone trunking formulas, should be used as an aid in selecting the specific system to be simulated, and also in evaluating the results of a simulation. The mathematical relationships are most useful in determining the size of trunk groups and of estimating the number of receivers required at a destination.

Simulation

Simulation is not new. Machines have been built to simulate specific telephone switching systems,[2] and devices such as a-c boards and network analyzers have been used to simulate electric power distribution systems. With the advent of large high-speed digital computers the cost of simulation is greatly reduced. The general principles of performing simulations on computers have been discussed in the literature.[3,4]

In the simulation of the operation of a data-message-switching system the computer calculates the characteristics of each message that would be offered to the system during the peak hour. It also calculates the progress that each message would make through the system and determines the delays it would encounter in waiting for the completion of preceding messages and in waiting for switching operations. Finally, the computer arranges the results in a form convenient for comparison with the results of other simulation runs.

The major steps in the computer program for the simulation are the determinations of the following:

1. Length of each message.
2. Time during the peak hour at which the message is available for transmission (offered time).
3. Originating source at which the message would be offered.
4. Destinations to which the message would be addressed.
5. Machine at the source which would first be available to transmit the message if there is more than one transmitter at the source.
6. Time at which each destination would have assigned an idle receiver.
7. Time at which idle receivers would have been connected and transmission would have started for a message sent simultaneously to several different receivers.
8. End of message time when connections would have been released.
9. Delays met by the message in waiting for other messages which precede it.
10. Delays imposed on the message at different points in the switching process.

PROGRAMMING THE COMPUTER

Because of their high cost and high speed, modern large-scale digital computers are usually operated by a trained staff capable of converting the programmer's instructions into punched cards or magnetic tape, running the problem on the machine, and returning the printed results to the engineer. They have automatic program routines which enable them to accept a program of instructions in the form of simple mnemonic codes each of which refers either to frequently used sequences of machine instructions or to a convenient group of storage registers. The programmer can also use relatively simple format instructions so that the printed results are in a form that is easy to evaluate.

It will be to the advantage of the engineer responsible for the simulation if he either learns to write the mnemonically coded simulation program himself or assigns one of his subordinates to do it. The preparation of the simulation program often leads to a better understanding of the operation of the system and may lead to suggested improvements even before the simulation is run. If it is necessary to employ the aid of a programming specialist, an engineer should work closely with him to insure that this opportunity for better understanding of the operation will not be lost.

To be useful for simulation of a data-switching system the computer should be of a type that can be easily programmed for the following operations by use of mnemonic codes or subroutines:

1. The usual arithmetic processes.
2. Computation of the exponential of a positive or negative argument.
3. Computation of the logarithm of any positive number to the Naperian base.
4. Moving to one of two or three groups of instructions in the simulation program after an arithmetic operation depending on whether the result is positive or negative or zero.
5. Discarding least significant or most significant digits by shifting.
6. Computation of computer-storage ad-

Vol.
Pt. 1
1959

Paper 59-890, recommended by the AIEE Data Communications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing April 28, 1959.

F. J. Gross is with the Bell Telephone Laboratories, Inc., New York, N. Y.

dresses in the computer and insertion of such computed addresses in specified instructions during the computing processes.

Random Numbers

Simulation of a data-switching system requires the frequent use of random numbers. For example, experience with existing teletypewriter data systems teaches us that messages offered to them occur at random times just as they do in telephone systems. Messages are distributed randomly among sources and also among destinations.

Random numbers are listed in tabular form in books by several authors[5,6] and it is possible to reduce them to punched cards or magnetic tape and store them in the computer. A much more convenient source is a series of numbers called pseudo-random numbers that can be calculated as needed by the computer with only three or four instructions and two or three storage registers.[7,8] To form a series of random numbers in a computer, one derives from an available formula, a multiplier, and a multiplicand of several digits each, and stores them in the computer. When the computer forms the product of these two numbers, the product is split and the least significant half can be used.

1. As one of a series of random numbers.
2. To replace the multiplicand before repeating the multiplication to form the next random number in the series.

The multiplicands constitute a series of several million different numbers that are just as useful for simulation as those found in published tables.

In the simulation each random number is used as though the decimal point were to the left, namely as a decimal less than one.

Message Characteristics

The simulation of a data-message-switching system requires that the characteristics of the messages be such that the simulation will represent as closely as possible the operation of the actual system.

MESSAGE LENGTH

Studies of large numbers of messages in teletypewriter systems have taught us that message lengths are distributed exponentially in accordance with the following equation:

$$NP(x)=NK\epsilon^{-cx} \qquad (1)$$

where

N = the total number of messages expected in the simulation
x = the length of each message usually expressed in words
c and K are determined so that the summation of $P(x)$ for all values of x is unity for a specified average message length

In a practical simulation all values of x and $NP(x)$ must be integral numbers. To accomplish this the following equations have been derived from equation 1 in a form convenient for use in programming a digital computer

$$x=\frac{ds+\ln N-\ln NA}{d} \qquad (2)$$

where

$$d=\frac{N\ln N-\ln(N!)}{N(\bar{x}-s)} \qquad (3)$$

and

$$N!\approx\sqrt{2\pi N}\left(\frac{N}{\epsilon}\right)^{N} \qquad (4)$$

$\bar{x}$ = average message length
s = shortest message length
A is varied from one to N, in increments of one, to give the N message lengths required

The values of $\bar{x}$ and s are known constants that depend on the nature of the business operation that originates the message. For example, for domestic air lines x is usually 20 to 30 words or 120 or 180 teletypewriter characters, but for automobile manufacturers the average is usually 50 to 60 words or 300 to 360 characters. In one industrial system the average is several hundred words.

Having assumed an appropriate value of $\bar{x}$, the value N will be found by dividing $\bar{x}$ into the total number of words or characters the system is designed to originate in the simulated time interval which is usually the busy hour.

MULTIPLICITY OF DESTINATIONS

In every data-switching system each message is addressed to at least one destination. However, some business operations require that a certain proportion of messages be sent simultaneously to a multiplicity of destinations. The number of destinations may be two, three, four, or more; up to 200 in some cases.

The system being simulated will have a specified proportion of system load assigned one destination, two destinations, and so on. A suitable device for determining destination multiplicity of a given message consists of several steps:

1. Tabulate the loads, in words, to be delivered in terms of one-destination, 2-destination, 3-destination, etc., multiplicity.
2. Assign a number to the given message by forming a product of a random number and the total system load in words.
3. Subtract successively from this product the one-destination, 2-destination, 3-destination loads until a negative or zero remainder is obtained.
4. Assign to the given message the first multiplicity for which a negative remainder was obtained.

This insures that the more predominant multiplicity will be more frequently chosen by the random process as the message characteristics are developed.

A tally of the load assigned to each multiplicity should be maintained. The computer can be programmed so that if the load assigned to a particular multiplicity is reached or exceeded during the simulation, then either the next higher or lower multiplicity is chosen.

ORIGINS AND DESTINATIONS

For purposes of simulation it is desirable to assign a unique number to each originating point in the simulated system. The use of the system will determine the proportion of the load to be assigned each originating point. Here again, the most suitable device for selecting the point of origin for a message is to develop the product of a random number and the system load, and successively subtract origin loads until the result is zero or negative. That origin is chosen in which the subtraction changed the algebraic sign of the results. Here again a tally of the load assigned each origin should be kept by the computer. As the problem nears completion the program should be such that loaded origins are passed over.

Destinations should be numbered and selected in a manner exactly similar to that for points of origin.

Only one origin will be chosen for each message, but the selection of a destination for a message should be repeated by the computer in accordance with the chosen degree multiplicity of address for that message.

OFFERED TIME

For purposes of simulation, offered time of a message is defined as the time at which the message is available in encoded form at the transmitter source. Experience has shown that messages occur individually and collectively at random during an interval like a busy hour. Two methods are available for determining the offered time of each of the N messages used in the simulation.

In the first method a pseudo-random number is computed for each message.

This number is treated as though the decimal point were at the left so that the random number lies between zero and one. It is multiplied by the simulation interval in seconds, and this product, rounded to the nearest 0.1 or 0.01 second, is used as the offered time of that message. When this method is used the length, origin, and destinations of every message must be calculated and sorted in ascending order of the offered time before the computer can proceed with the simulation.

In the second method the offered time of any message is determined by calculating the elasped time from the offered time of the previous message. To do this a unit time interval is assumed which is short compared to the average interval between offered times. The number of times (M_1) that the elapsed interval between two successive offered times will be less than one unit time interval is given by

$$M_1 = N(1-\epsilon^{-\mu}) \tag{5}$$

where μ is the average number of messages per unit time interval. The number of times, M_T, that are less than T but more than T-1 unit time intervals will occur between offered times is given by

$$M_T = M_1\epsilon^{-T\mu} \tag{6}$$

The computer is programmed to prepare a table of rounded-off values of M_T for various values of T beginning with $T=1$. The computation of the table is stopped when the sum of all values of M_T is equal to N.

Since zero is used as an idle indication it was convenient to assume that the first offered time occurred at the end of the first unit time interval. For each subsequent offered time the computer is programmed to select a random number, multiply it by N, and successively subtract from the product the values of M_T in the table beginning with the table entry for unit time interval. The value of T for which the difference is zero or negative is the one to be used in determining the next offered time. This method provides a random entry into the table. A tally of the use of each value of T should be kept separately from the table and the computer should be programmed to pass over a value of T as the number of times it has been used equals its values of M_T. If the value $T-0.5$ rather than T is added to the previous offered time in computing the next, the last message will occur almost exactly at the end of the simulation time. This is a consequence of integrating equation 7 from $T-1$ to T.

The latter method was developed from the following equation which relates μ, the average messages per unit time interval, to the probability that the next message will be offered to the system T unit time intervals after the last message was offered. It is taken from reference 4.

$$P(T)dT = \mu\epsilon^{-\mu T}dT \tag{7}$$

When the results of the two methods were compared for a specific example, the distribution of messages per subinterval for the second method agreed with the Poisson distribution nearly as well as distribution for the first method.

The first method of computing the offered time has been used on a medium-size computer of 2,000-word internal storage. In this case message characteristics can be punched on cards in one computer run, the cards being passed through an off-line sorter to arrange them in ascending order of offered time and then fed to the computer as needed during the second or simulation run.

The second method has been used on a large computer of 32,000-word storage which was capable of doing the entire problem in one run. This method avoids sorting which would require profligate use of expensive computer time.

Simulation of the System

Having programmed the computer to produce message characteristics in a form suitable for use as inputs to the simulation, consider the systems to be simulated.

Each automatic data-message-switching system comprises one or more of the following major elements: message source, message transmitter, common control, destination, and receivers.

A message source is a point in the system where a data message becomes available for transmission and where it can be held or stored until the switching system is ready to transmit it. Each source is served by one or more transmitters or transmission paths.

The primary function of common control is to sequence the bidding of transmitters for destinations so that one receiver will not be simultaneously connected to two transmitters.

Destinations are points in the system to which data messages are addressed and to which transmitters seek a connection before transmitting the message. Each destination is served by one or more receivers.

In some systems a receiver for one transmission link in the system may serve as a source for another transmission path which will relay the message on toward its destination. To avoid ambiguity, the source where the message first becomes available to the switching system will be referred to as an originating source.

In a real system each of these elements has one of the three following states or conditions:

1. Idle.
2. Busy performing its function with a calculated future completion time.
3. Seized waiting for some switching function to complete or for a busy unit to become available.

In simulation on a digital computer it has been found convenient to use a single storage register to remember the state of each switching element. In each such register the idle condition is represented by zero, the busy condition is represented by the time at which the condition is to end, and the seized condition is represented by a very large constant. If this is done, then proper programming will enable the computer to test the state of the simulated element by testing the contents of the storage register for zero, or by subtracting the contents of the simulation time register from the contents of the state register and testing the result for positive, negative, or zero.

A single storage register is normally sufficient to remember current simulation time during the computer run. At the start of the simulation this register is set to the unit time interval used in the simulation, usually 0.1 or 0.01 second. Although the most obvious manner of conducting the simulation might be to advance the simulation time in small steps and examine all state registers at each step for a change from busy to available, this method is wasteful of computer time. A better method is to program the computer to determine the earliest change of state anywhere in the system and to advance the simulation time register to that time. The computer should be programmed so that after it processes all changes of state occurring at that simulation time it determines the time of the next earliest change of state in the system and advances the simulation time accordingly. In a completely idle system the simulation time is advanced to the offered time of the next earliest message.

For those classes of system elements in which each element has the exclusive use of one or more transmission paths, computer registers are provided for each element to store the addresses of the state registers of the paths used by that element. For example, if some originating sources are served by more than one transmitter than appropriate registers in the computer should be designated for each source

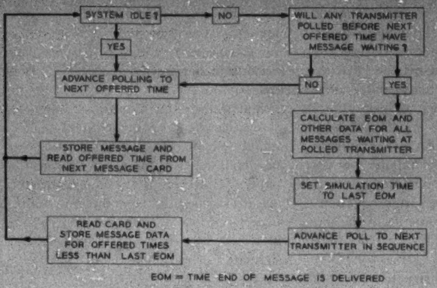

SYSTEM IDLE ?
NO
WILL ANY TRANSMITTER POLLED BEFORE NEXT OFFERED TIME HAVE MESSAGE WAITING ?
YES
ADVANCE POLLING TO NEXT OFFERED TIME
NO
YES
CALCULATE EOM AND OTHER DATA FOR ALL MESSAGES WAITING AT POLLED TRANSMITTER
STORE MESSAGE AND READ OFFERED TIME FROM NEXT MESSAGE CARD
SET SIMULATION TIME TO LAST EOM
READ CARD AND STORE MESSAGE DATA FOR OFFERED TIMES LESS THAN LAST EOM
ADVANCE POLL TO NEXT TRANSMITTER IN SEQUENCE
EOM = TIME END OF MESSAGE IS DELIVERED

Fig. 1. The flow diagram for the simulation of a single-line selective-calling data-message-switching system

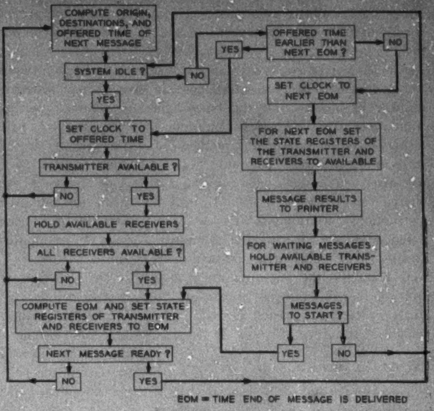

COMPUTE ORIGIN, DESTINATIONS, AND OFFERED TIME OF NEXT MESSAGE
OFFERED TIME EARLIER THAN NEXT EOM ?
YES
NO
SYSTEM IDLE ?
NO
SET CLOCK TO NEXT EOM
YES
SET CLOCK TO OFFERED TIME
FOR NEXT EOM SET THE STATE REGISTERS OF THE TRANSMITTER AND RECEIVERS TO AVAILABLE
TRANSMITTER AVAILABLE ?
NO
YES
MESSAGE RESULTS TO PRINTER
HOLD AVAILABLE RECEIVERS
ALL RECEIVERS AVAILABLE ?
FOR WAITING MESSAGES HOLD AVAILABLE TRANSMITTER AND RECEIVERS
NO
YES
COMPUTE EOM AND SET STATE REGISTERS OF TRANSMITTER AND RECEIVERS TO EOM
MESSAGES TO START ?
NEXT MESSAGE READY ?
YES
NO
NO
YES
EOM = TIME END OF MESSAGE IS DELIVERED

Fig. 2 (left). The flow diagram for simulation of a data-switching office with negligible switching time per message

and the addresses of the state registers for the transmitters serving that source should be stored in them. Usually destinations will be provided with similar registers for addresses of state registers for receivers.

In addition to the digital computer storage registers already described, it has been found desirable to provide a large block of registers for the active messages. A message is considered active when its offered time becomes equal to, or less than, the current system simulation time. The messages are sequenced in order of offered time. As a message becomes active it is assigned sufficient registers for the message characteristics previously described, for the address of the state registers of each transmitter and receiver that will be assigned to it, for the time that each transmitter and receiver will become available to the message, and for the times at which transmission starts and ends. In the time-of-availability registers in the message storage block, a zero should indicate that the element is not yet available to the message. The contents of some of these registers are used to calculate message delays.

When the time of the end-of-transmission equals the system simulation time, the computer program should be such that the message is removed from the active storage and the various delays computed and combined with results of previous messages as required by the problem.

Finally, if the common control either sequences or polls transmitters in a particular sequential order then a storage register should be provided to store the identity of the latest transmitter polled. An additional storage register may be needed to remember the simulation time at which the latest transmitter has been polled.

FIRST EXAMPLE

The simulation of a single-line teletypewriter selective calling system is an example of a small system that has been simulated on a moderate-size magnetic-drum digital computer of 2,000 10-digit-word capacity.

The system to be simulated involved half-duplex operation of a single line to which is connected a multiplicity of receiving units and transmitting units. A common control or master control installed at one of the transmitter sites polls the transmitters in a fixed sequence after any station completes its transmission. Polling is continuous when the system is idle. When a transmitter has messages to transmit, it begins transmission only when it is polled by the master control unit. The transmitter is assumed to transmit all messages on hand when polled.

Since only one message could occupy the line at any time and since receivers are connected on the fly in response to address codes, the destination and multiplicity of addresses did not affect the simulation results and were not considered. All other message characteristics were calculated as previously described.

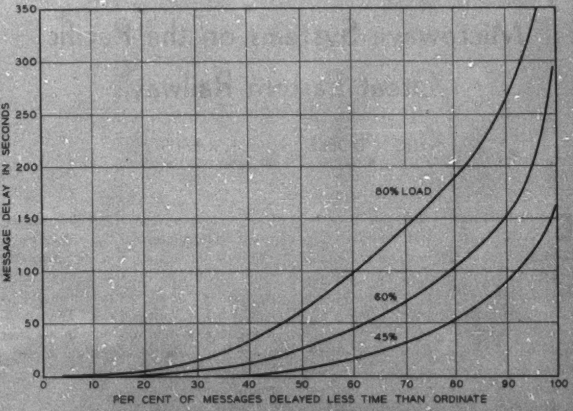


Fig. 3. Cumulative probability of delay from the simulation of a single-line selective-calling system

Transmitter loads were assumed such that their sum did not exceed the desired loading of the line. The load, transmission speed, and average length of message were assumed and this determined the number of messages in the busy hour simulated. Characteristics of every message were computed and punched on cards in one run on the computer. These message cards were sorted in order of offered time on a sorter. A program of instructions conforming to Fig. 1 was prepared and loaded into the computer. During the simulation the message cards were read into the computer under control of the program as illustrated in Fig. 1. Results for each message were punched on cards.

SECOND EXAMPLE

The simulation of a multiline-data-switching system with one switching center is an example of a large system that has been simulated on a high-speed computer with 32,000 words of core storage and on-line tape storage.

From the system to be simulated the transmission speed, average message length, load per source, load per destination, transmitters per source, and receivers per destination were determined. The switching center was assumed ideal with negligible connection time and an infinitely fast sequence circuit in common control.

Offered times were to be computed as needed with 0.1 second minimum between any two. With this ideal arrangement delays would be caused only by messages waiting for previous messages. Results were to be compared later with those of an actual system to see how closely the actual system approached the ideal.

Message lengths were calculated by the computer in accordance with equation 2. With the aid of pseudo-random numbers the lengths were distributed randomly throughout a block of computer storage.

Offered times were computed as needed using the table derived from equation 6 and the other characteristics of the message were computed along with the offered time.

The program of instructions for the system simulation was prepared in conformance with Fig. 2. Computer registers assigned to the sources and destinations contained the addresses of state registers of transmitters and receivers allotted them.

Results

The computer can be programmed to calculate average delays imposed by different parts of the switching systems by summing the delays and dividing by the number of messages. It can be also programmed to maintain appropriate tallies so that cumulative probability curves like those of Fig. 3 can be quickly plotted from printed computer results.

Conclusions

Simulation on a digital computer is a powerful tool for comparing the capacity and speed of delivery of different data-switching systems or of different systems layouts of one type of switching system for given load requirements, before committing money for major design effort or construction.

Simulation of larger systems than those shown can be programmed. Also, more minute details of system operation can be simulated than those described in this paper. However, if the computer storage is exceeded or the computer running time is excessive, then the system simulated is probably too large or it is being simulated in too great detail.

References

1. A FULL AUTOMATIC PRIVATE-LINE TELETYPEWRITER SWITCHING SYSTEM, **W. M. Bacon, G. A. Locke.** *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 473–80.

2. A THROWDOWN MACHINE FOR TELEPHONE TRAFFIC STUDIES, **G. R. Frost, W. Keister, A. E. Ritchie.** *Bell System Technical Journal*, New York, N. Y., vol. 32, 1953, pp. 292–359.

3. SIMULATION—ITS PLACE IN SYSTEM DESIGN, **H. H. Goode.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 39, 1951, pp. 1501–06.

4. SYSTEM ENGINEERING (book), **H. H. Goode, R. E. Machol.** McGraw-Hill Book Company, Inc., New York, N. Y., 1957, p. 96, pp. 403–07.

5. A MILLION RANDOM DIGITS (book), **RAND Corporation.** The Free Press, Glencoe, Ill., 1955.

6. TABLE OF RANDOM SAMPLING NUMBERS (book), **M. G. Kendall, B. Babington Smith.** Cambridge University Press, London, England, 1946.

7. GENERATING AND TESTING OF PSEUDO-RANDOM NUMBERS, **O. Taussky, J. Ford.** "Symposium on Monte Carlo Methods" (book), **H. A. Meyer,** editor. John Wiley & Sons, Inc., New York, N. Y., 1956, pp. 15–28.

8. THE GENERATION OF PSEUDO RANDOM NUMBERS ON A DECIMAL CALCULATOR, **Jack Moshman.** *Journal*, Association of Computing Machinery, Pittsburgh, Pa., vol. 1, no. 2, Apr. 1954, pp. 88–91.

Microwave Systems on the Pacific Great Eastern Railway

R. N. DOBLE
NONMEMBER AIEE

J. E. RAFTIS
ASSOCIATE MEMBER AIEE

EXISTING microwave systems on the Pacific Great Eastern Railway have been depicted on the railway route map. See Fig. 1.

The company provides freight and passenger service over an 800-mile single-track railway system which extends north and south through the interior of the Province of British Columbia in Canada. Its right of way traverses the Coastal, Selkirk, and Rocky Mountain ranges. The company operates its communication facilities for train service and, in addition, leases telephone and telegraph channels to both common carriers and private users. Prior to the installation of microwave radio facilities, communications were provided over open-wire pole lines and via hand-written train orders. That communication system was only in use over about 50% of the present railway route.

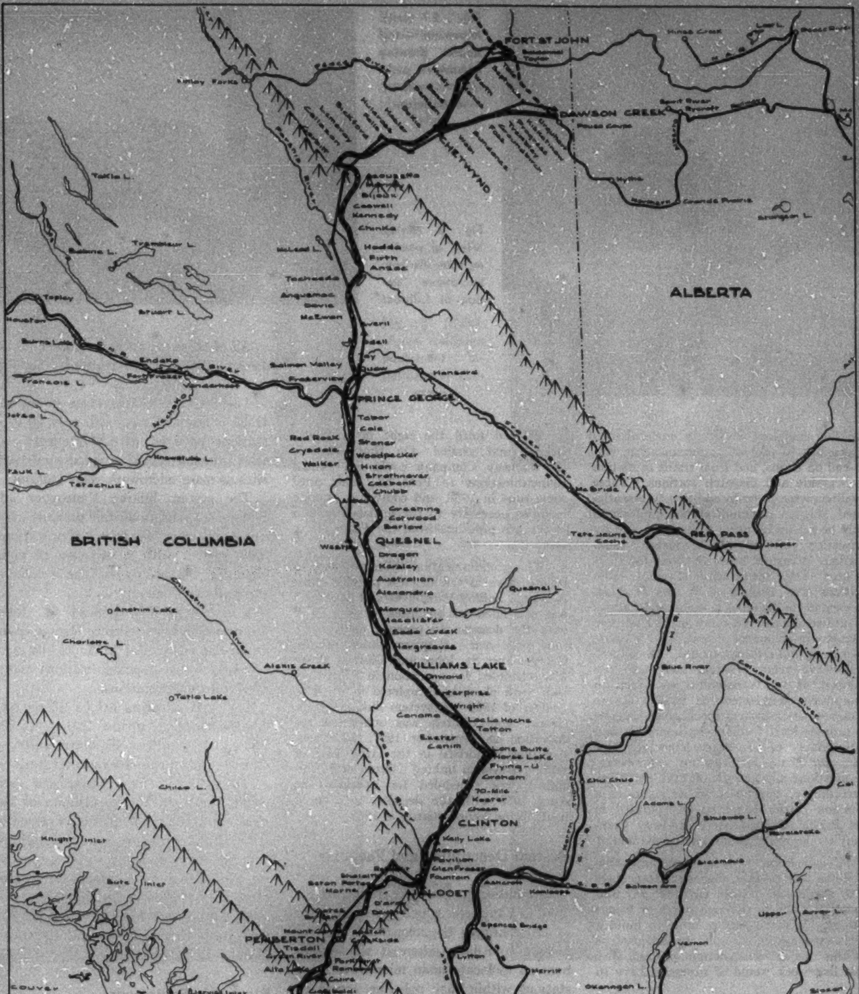
Planning

When Pacific Great Eastern contemplated extending its railway lines geographical, operational, and scheduling requirements indicated the installation of comprehensive radio systems.

1. Over most of the planned railway route, terrain was rugged and mountainous: There were areas subject to landslides, there were sections of heavy timber, and there were large stretches of swampy ground. These natural obstacles imposed high initial costs on the construction of a pole line, and subsequently expensive maintenance in its operation. A further consideration was that wire lines might not prove dependable enough for the type of railway service planned.

Paper 59-889, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing April 24, 1959.

R. N. DOBLE is with the Pacific Great Eastern Railway Company, Vancouver, B. C., Canada. J. E. RAFTIS is with Canadian Motorola Electronics Ltd., Toronto, Ont., Canada.



FORT ST JOHN
Kiskatinaw
Taylor
Kelley Forks
Hines Creek
Peace River
DAWSON CREEK
Spirit River
Sexsmith
Pouce Coupe
Hythe
Wembley
Grande Prairie
Sturgeon L.
CHETWYND
Azouzetta
Kennedy
Chuka
Hodda
Firth
Anzac
McLeod L.
Tacheeda
Anguemac
Davie
McEwan
Takla L.
Trembleur L.
Babine L.
Stuart L.
Topley
Houston
Burns Lake
Endako
Fort Fraser
Vanderhoof
Francois L.
Salmon Valley
Fraserview
Hansard
PRINCE GEORGE
ALBERTA
Tabor
Cale
Stoner
Red Rock
Woodpecker
Hixon
Strathnaver
Walker
Chubb
Tetachuk L.
McBride
Fraser River
Tete Jaune Cache
Jasper
BRITISH COLUMBIA
QUESNEL
Dragon
Australian
Alexandria
Marguerite
Macalister
Soda Creek
Hargreaves
Quesnel L.
Anahim Lake
Charlotte L.
Chilcotin River
Alexis Creek
Tatla Lake
WILLIAMS LAKE
Blue River
Columbia River
Enterprise
Wright
Lac La Hache
Exeter
Canim
Lone Butte
Horse Lake
Flying-U
Graham
70-Mile
Chasm
CLINTON
Chilco L.
Chu Chua
North Thompson
Adams L.
Shuswap L.
Revelstoke
Knight Inlet
Bute Inlet
Kelly Lake
Pavilion
Fountain
Seton Portage
LILLOOET
Ashcroft
Kamloops
Salmon Arm
Sicamous
Upper Arrow L.
Spences Bridge
PEMBERTON
Lytton
Merritt
Vernon
Okanagan L.
Kelowna
Brookmere
Slocan
North Bend
Boston Bar
Courtenay
Powell River
Jervis Inlet
SQUAMISH
Britannia
Brunswick
Island
Alberni
Port Alberni
Parksville
Nanaimo
Horseshoe Bay
NORTH VANCOUVER
VANCOUVER
Hope
Princeton
Penticton
Skaha L.
Ladysmith
Lake Cowichan
CANADA
U.S.A.
Bellingham
Port Renfrew
VICTORIA
Mount Vernon
Rockport

Fig. 1. Route map of railway and radio systems

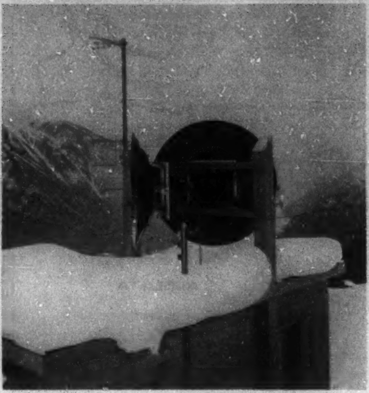


Fig. 2 (left). Mountain - top station showing microwave paths with railway route below

Fig. 3 (right). View of passive repeater site from microwave station at Lillooet

2. The company wished to modernize its operations to provide communication end-to-end on trains, and from trains in motion to wayside and dispatch stations. These facilities were to be in addition to essential fixed station communications, and could only be set up by using train radio, and by providing vhf (very-high-frequency) radio stations at regular intervals along the right of way. Once the principle of remote radio stations was established it was practical to expand the facilities and choose locations to utilize microwave relay in lieu of pole lines to interconnect dispatchers, wayside operators, and railway termini. Once these microwave relay stations had been developed, it was economical to select an equipment complement to include also a communication system specifically designed to provide lease facilities for operation and maintenance of the Peace River gas pipe line, and lease facilities for other communications users south of Fort St. John. Accordingly, assertation of its rights as a common carrier could be incorporated in the Pacific Great Eastern Railway Company's microwave radio system planning.

3. Construction of the communication facilities had to take place simultaneously with survey of railway right of way and laying of track in the area north of Prince George. While it was possible to build a microwave system completely independent of this major construction project, any pole line work would of necessity have to be delayed until the right of way was cleared and graded. On that schedule the Railway Company could not extend communications to Fort St. John until some time in 1958, and could not provide facilities necessary to the initial operation of the gas pipe line from the Peace River area.

Table I. Traffic Channels

Facility	Fort St. John and Prince George	Prince George and Clinton	Clinton and Vancouver
Voice channels	32	18	60
Remote control channels	2	2	Nil
Teletype channels	20	12	14
Telemetering channels	future requirement		
Vhf radio control channels	9	5	7
Supervisory channels	5	5	5

4. The communication systems were to be operative in the shortest possible time. A system linking Vancouver and Squamish was scheduled for operation in June 1956 with the decision to proceed made early that same year. A system linking Prince George, Fort St. John, and Dawson Creek was scheduled for operation in September 1957 with planning completed in the first quarter of 1956. A system linking Prince George and Clinton was scheduled for operation in September 1957, with construction authorized in the latter part of 1956. A system linking Clinton and Vancouver was scheduled for operation in April 1959 with the decision to proceed made in the second quarter of 1958.

System Description and Facilities

In addition to the links from Fort St. John to Vancouver which have a capacity of 120 channels, the microwave systems include spur links where it was uneconomical to locate main microwave relay stations within easy pole-line access to railway stations. Performance requirements for the systems are stringent to accommodate railway operations without benefit of backup communication services. Major items of electronic equipment comprise a comprehensive and complex system of communications.

1. The longest microwave path on the system is over 60 miles.

2. The shortest microwave path on the system is under 10 miles.

3. The average microwave path on the system is about 25 miles.

All of these distances are measured between active repeater stations. See Fig. 2.

The 800-mile vhf coverage of railway track is accomplished with only 20 drop stations powered with 2-frequency 60-watt transmitters utilizing the same radio sites as main microwave relay stations.

The system linking Vancouver and Squamish spans an air-line distance of 36 miles and comprises four active stations. This system with a capacity of eight channels, was the first trial use of radio by the Railway Company.

The system linking Fort St. John, Dawson Creek, and Prince George spans an air-line distance of 258 miles and comprises 11 active repeater stations and 5 passive repeater stations.

The system linking Prince George and Clinton spans an air-line distance of 222 miles and comprises six active repeater stations and two passive repeater stations.

The system linking Clinton and Vancouver spans an air-line distance of 180 miles, and comprises 11 active repeater stations and 8 passive repeater stations. This system was laid out without employing towers to elevate antennas at microwave relay stations. See Fig. 3.

As authorized for construction, the main microwave links included the traffic channels in Table I.

Electronic Equipment

Microwave r-f (radio-frequency) equipment is of Motorola design and operates in the 6,000-mc (megacycle) frequency band. System requirements pertaining to propagation reliability are met through use of frequency diversity. The r-f baseband has a bandwidth of 1 mc. See Fig. 4.

Voice multiplex equipment is of Philips design and is of the single-sideband type. Two completely independent voice multiplex systems are in use. One system

Vol.
Pt. 1
195

Fig. 4 (left). Equipment racks showing microwave control panels and multiplex junction hybrid

Fig. 5 (right). Voice multiplex equipment installation at Vancouver

Fig. 6 (left). Party-line dial panels telegraph loops, vhf drop radio, system interrogation alarm, and microwave for spur link

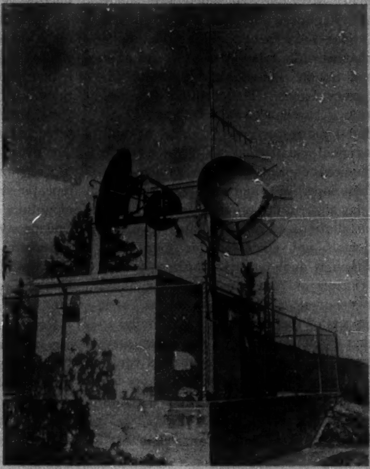


Fig. 7 (right). Microwave antenna systems for both main and spur links at remote site

provides drop and insert circuits at each relay station, and has an ultimate capacity of 16 channels. The other system provides trunk circuits at selected relay stations and has an ultimate capacity of 120 channels. Submultiplex equipment for party-line teletypewriter channels and for alarm, control and telemeter circuits is of Motorola design. This equipment, which is available on a drop and insert basis at each relay station, provides up to 18 channels which can be applied to any carrier derived voice channel. Submultiplex equipment for trunk teletypewriter channels is of Philips design. This equipment which terminates at Fort St. John, Prince George, and Vancouver will provide up to 18 teletypewriter channels over selected carrier derived voice channels. See Fig. 5.

Motorola designed vhf drop radio equipment operates in the 150-mc band. Not all microwave relay stations are equipped with a vhf drop station, as the criteria for selection of suitable vhf sites was complete coverage of the railway right of way with minimum signal overlap between adjacent stations. See Fig. 6.

Auxiliary equipment consists of a Motorola remote supervision and control system working on the interrogation principle, providing five condition reports for each microwave relay station; a Motorola vibracode selective-calling system from Pacific Great Eastern dispatch offices to trains and wayside stations; separate party-line 2-digit dial systems to link wayside stations on the railway and

Fig. 8 (left). View of building and antenna system at Lone Butte

Fig. 9 (right). View of passive repeater site from microwave station at Bowen Island

to link pumping stations on the pipe line. Party-line dial panels were a part of system design for the project.

Spur links comprise Motorola multichannel radio relay systems operating in the 150-mc, 960-mc, and 6,000-mc frequency bands. These equipments provide a basebandwidth of 24 kc, 42 kc, and 1 mc respectively. Single- or dual-terminal installations are dictated by the required dependability in service. Remote supervision and condition reports on relay stations are provided in some of the installations. Frequency diversity operation is not employed on spur links, although dual terminals operate on the principle of "hot standby" to minimize switchover time. Multiplex on spur links is of the short-haul type, similar to that employed for drop and insert service on main links. Submultiplex equipment uses a frequency-shift voice-frequency carrier. See Fig. 7.

Nonelectronic Equipment

Antenna systems at the microwave relay stations usually consist of a tower-mounted reflector and parabolic dish combination, with associated short waveguide runs. Directional antenna arrays are used at vhf drop stations and are located below the reflectors on microwave towers. See Fig. 8.

Passive repeaters comprise both self-supporting and guyed structures. Reflecting surfaces range from 640 square feet to 96 square feet under control of the path lengths and the angle of interception of the microwave beam. As many as three passive reflectors are required to establish the r-f path in several of the microwave hops. See Figs. 9 and 10.

One of the most interesting, and from a technical standpoint one of the most challenging, passive repeater installations is that in the Pine Pass area between stations at Azouzetta Lake and Nelson Creek. Here the microwave beam is actually subjected to a double reversal while passing over the natural obstacle of the Murray mountain range. The passive repeater stations are located on adjacent mountain peaks about a mile apart. One passive repeater looks down at the Azouzetta site from about 1.5 miles away, while the other faces the site on Nelson Creek about 20 miles away. See Fig. 11.

Remote microwave relay stations make use of a variety of a-c power services. Where possible, commercial power is used with stand-by diesel power plants. In some areas twin-diesel engines were installed to provide both main and stand-by service. These installations include a 6,000-gallon storage tank to provide fuel through the long winter period when the station may be inaccessible to heavy vehicles. In other areas the main power is provided by natural gas engine generators, and propane gas storage tanks were installed for stand-by service in the event of a pipe-line failure. See Fig. 12.

Capacity of stand-by power plants ranges from 7.5 to 15 kw. All plants, regardless of fuel type and capacity, were purchased from one supplier to ensure uniformity in system design and interchangeability of parts within practical limits.

Buildings are of concrete block construction with a poured concrete roof, and with projecting supports for mounting parabolic antennas. Size of equipment rooms ranges from 250 square feet to 150 square feet, while size of generator rooms ranges from 150 square feet to 100 square feet. Equipment rooms contain thermostatically controlled electric heating and ventilation, and are designed to accommodate racks up to 9 feet in height, of either the self-supporting or wall-mounted type.

Performance

Each of the three separate sections which, when combined in tandem, make up a continuous radio baseband from Vancouver in the south to Fort St. John in the north, have similar operational requirements.

1. The radio baseband must accommodate 120 voice channels derived by single-side-

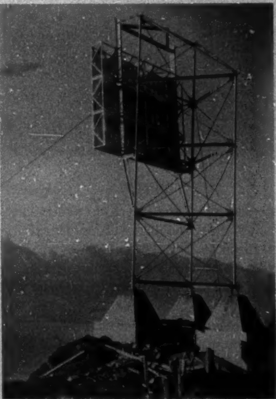


Fig. 10. Passive repeater in Vancouver to Bowen Island path

Vol.
Pt. 1
1959

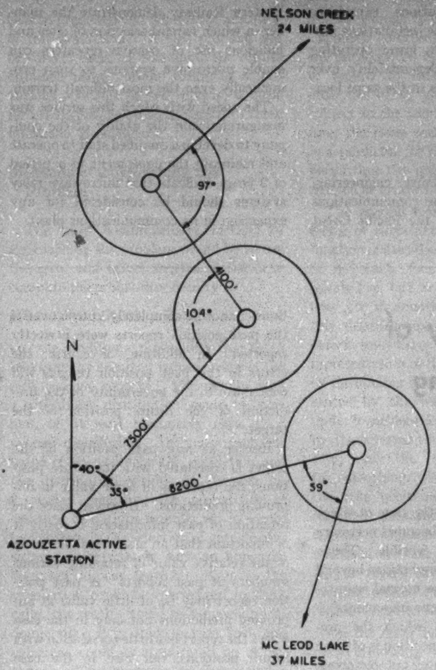


Fig. 11 (left). Microwave paths in the Pine Pass area

Fig. 12 (above). Stand-by engine generator installation at Vancouver

band multiplex equipment conforming to International Telephone Consultative Committee recommendations.

2. Additional channels for Teletype shall be derived by means of submultiplex equipment operating on the frequency shift principle, and conforming to International Telegraph Consultative Committee recommendations.

3. Propagation reliability shall be such that the r-f carrier-to-noise ratio shall exceed 30 db (decibels) for 99.96% of the time averaged over one year.

4. Over any one section, the noise level from all sources on the worst voice channel during periods of normal traffic, shall not exceed 30 dba (decibels adjusted) referred to 0 dbm (decibels referred to 1 milliwatt) transmission level.

Although the microwave baseband will accommodate 240 voice channels, appreciable savings in antenna and passive repeater costs are afforded by designing the system for the transmission performance and channel loading as specified. The average r-f carrier-to-noise ratio per hop is 38 db. This figure in combination with the microwave and multiplex used, has been demonstrated by field test, and has been proved by operating experience to meet the transmission objectives of the company, and to provide an adequate operating margin.

Operation

In 1956, when the Pacific Great Eastern Railway Company decided to begin the construction and operation of these complex microwave communication facilities, its communications staff comprised only one person with a background of knowledge and experience in radio communications.

The company added some eight maintenance technicians and a supervisor during the period to mid-1957, whose responsibilities were to operate and maintain microwave stations spanning over 600 miles of rail line. These persons were given a course of instruction covering all major equipment to be installed. At the end of the course a written examination was given, and the results were used by the chief communications officer at Pacific Great Eastern Railway Company to assist in assigning staff responsibilities. Subsequently the new employees assisted in the check-out of equipment and in system line-up.

The company employed four additional maintenance technicians during 1958 to maintain the 200-mile section south from Clinton. As with their predecessors, these persons were carefully trained prior to and during installations. While the operating group was being built up, the company also expanded its communications engineering and administration group to number six persons, and established a central maintenance depot in Vancouver.

A special control feature with which vhf drop stations are equipped greatly simplifies operational procedure on train-to-dispatcher traffic. All trains in motion are connected to the dispatcher via a party-line telephone circuit. Similarly all wayside stations are connected to the dispatcher via a separate party-line telephone circuit. All transmitters at vhf drop stations are turned off, while all receivers are turned on. When a train originates a call to the dispatch office, the signal is applied to the party line at only one vhf drop station where the vhf transmitter connecting to the train is automatically turned on. The transmitter remains operative for about 10 seconds after completion of the message. The train crew is then immediately advised if the dispatcher is in communication with another train and cannot engage in conversation. Alternately, if the call requires acknowledgment only, this may be given by the dispatcher without the need to select the proper vhf drop station along the route.

All microwave and vhf drop repeaters are equipped with a squelch circuit which prevents noise output from the receiver from loading the microwave baseband in case of path or equipment failure. This device has made significant improvements in the over-all dependability of the communications system in two ways:

1. Without vhf receiver squelch, an entire microwave channel can become saturated by noise in the event of an inconsequential vhf link failure.

2. With microwave receiver squelch, reliable communications can be established on either side of a major system break.

Although the passive repeaters, which are located at altitudes up to 7,000 feet above sea level, are subject to extreme climatic conditions, including winds of hurricane violence and ice deposited by low-level cloud formations, experience has indicated that these installations provide a lower first cost, lower operating cost, and improved dependability over active repeater stations in the same locations.

Conclusions

The successful planning, engineering, and construction of the communications facilities described for the Pacific Great Eastern Railway demonstrate the manner in which careful selection of sites and judicious use of passive repeaters can enable microwave systems to span economically even the most difficult terrain.

The speed with which this service was inaugurated and the ability of the company to develop a qualified staff to operate and maintain the equipment in a period of 2 years, indicate that microwave relay systems should be considered for any expansion in its communications plant.

An Approach to the Study of Automatic Target Tracking

LEONARD PODE
NONMEMBER AIEE

SEVERAL problems encountered in the design of automatic processing systems for data obtained from search radars give rise to a requirement for the prediction of the present or future position of a target on the basis of past position reports. The most obvious of such problems is that of scan-to-scan correlation of target reports. Here the prediction of the present position of the target determines the center of the "gate" in which the target is sought; and the accuracy of this prediction determines the size of the gate. It is important that the prediction be accurate and the gate size reasonably small in order to avoid confusion between neighboring targets. Any scan-to-scan correlation system will break down when the density of targets becomes too great. The accuracy of prediction determines the density at which trouble may be encountered. While it is not expected that target densities will ordinarily attain unmanageable proportions, this situation could easily arise due to enemy action in launching decoys and chaff.

The need to maintain the integrity of target tracks as long as possible in the event that direct sensing of the target position is temporarily lost also makes demands on the accuracy of prediction. Such a situation may arise in the presence of electronic countermeasures or because of other contingencies that may interrupt the normal functioning of the radar system. Of course, when the interruption of target-sensing services is prolonged, past information will become useless and existing tracks would have to be dropped; but it is desirable that the time at which the dropping of targets becomes necessary be extended as far as feasible. Therefore, the extension of predictions beyond what is required for the normal scan-to-scan correlation is of some importance.

Another problem in which the prediction of future target position is of some importance is that of intercept vectoring. Although in this case tactical considerations would enter, predictions formulated without regard to tactical factors, such as will be considered in the present discussion, should be of some interest.

The factors limiting the accuracy with which the future position of the target can be predicted are; first, the maneuverability of the target, and second, the accuracy of past position reports. The possibility of predicting the future position of the target derives from the fact that the target does not have infinite maneuvering capabilities. Its velocity is finite and it has very limited ability to change the magnitude of its velocity. Although it has the capability of orienting its velocity in any direction, the rate at which the direction of velocity can be changed is also limited. Maneuvers are not only limited physically, but also by choice, inasmuch as the target suffers in the attainment of its objective if it is constantly turning. Because the maneuverability of the target is not infinite there will be some correlation between the past positions of the target and its future course. On the other hand, because the target is maneuverable, this correlation will not be perfect so that its future positions cannot be completely known even if the past position reports were perfectly reported. In addition, of course, the errors in the past position reports will contribute to the uncertainty of the prediction of the future position of the target.

Insofar as any past position of the target is correlated with its future positions, the report is of some value in improving predictions. However, since the retention of past information is costly it is important that an assessment be made of the relative value of retaining various numbers of past reports A past position report may be of little value in improving predictions not only in the case where the report has little correlation with future positions, but also in the case where the report is very highly correlated with another report that is retained. Thus, not only the number but the distribution in time of the retained reports is important. The latter circumstance is particularly likely to arise when the system is not synchronous as, for example, when tracking is to be accomplished on the basis of returns from multiple radars operating at different scanning rates.

To facilitate the examination of questions regarding the number and distribution of points to be retained as well as to form a basis for the selection of a prediction method, a solution of the following problem is sought: Given the past position reports $\mathbf{R}_1$, $\mathbf{R}_2$ ..., $\mathbf{R}_n$ at times $t_1, t_2, \ldots, t_n$ respectively, t_1 referring to the most recent report, in consideration of the statistics of the errors associated with these reports and the statistics of target maneuvers, how can the position of the target, $\mathbf{R}_0$, at the future time, t_0,

---

Paper 59-925, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21-26, 1959. Manuscript submitted October 21, 1958; made available for printing May 28, 1959.

LEONARD PODE resides in Sherman Oaks, Calif.

Vol.
Pt. 1
1959

best be predicted and what is the accuracy of this prediction? If this problem can be solved, then, by examining the results obtained using various values of n and various spacings of the reports, a basis for intelligent decisions in regard to these questions is achieved.

Methods of Approaching the Problem

The most comprehensive method of approaching the problem would be to use the error and target motion statistics to compute the probability density

$$p[\mathbf{R}_0(t_0)|\mathbf{R}_1(t_1), \mathbf{R}_2(t_2), \ldots \mathbf{R}_m(t_m)]$$

i.e., the conditional probability that $\mathbf{R}_0(t_0)$ takes on any specified value in view of the past position reports. Presumably such a probability distribution would be normal, or at least reasonably close to a normal distribution; the most probable value of $\mathbf{R}_0$ would be the prediction and the distribution would also furnish information regarding the probability of specified deviations from $\mathbf{R}_0$ which would be valuable in the selection of gate dimensions. Unfortunately such an approach does not appear feasible. Even if the calculation could be carried out analytically, the results would be somewhat academic since it is unlikely that such calculations could be duplicated on a real time basis by a computer of reasonable size.

The compromise that is usually made in such a situation is to use a linear predictor and adjust the constants by the method of least squares. Thus the predicted value of $\mathbf{R}_0$ is taken to be

$$\mathbf{R}_0 \text{ est} = a_1\mathbf{R}_1 + a_2\mathbf{R}_2 + \ldots + a_n\mathbf{R}_n \quad (1)$$

where $a_1, a_2 \ldots, a_n$ are to be adjusted to minimize the expected or average value of $|\mathbf{R}_0 - \mathbf{R}_0 \text{ est}|^2$. The justification for this method of prediction is partly that it can be expected to be accomplished by a realizable computer, partly that it is reasonably amenable to analysis, and that, in addition, though surely not completely general, it is of fairly broad generality. An idea of the generality of the linear predictor can be gained from the consideration that this is the form that would result if the co-ordinates of $\mathbf{R}_1$, $\mathbf{R}_2$ $\ldots$, $\mathbf{R}_n$ are fitted by a linear sum of selected functions of time. These may be a set of polynomials in t, or a set of sinusoids, or any arbitrary set of functions of t.

It is assumed that the target height does not change to any significant extent so that the motion is essentially two-dimensional and that the range of interest is small enough to permit the use of plane Cartesian co-ordinates. Since, for present purposes, no distinction will be made in regard to directions, the two co-ordinates of x_i, y_i of any report $\mathbf{R}_i$ are subject to the same statistics and, therefore, the same coefficients would appear in a predictor for either co-ordinate, thus permitting the predictor equation to be written in vector form as in equation 1.

The encompassing by the linear predictor of curve-fitting with selected time functions raises the question as to whether or not such time functions should be selected or left arbitrary as, in essence, they are in equation 1. The answer to this depends upon whether there is *a priori* knowledge of a time function representation of the motion of the target. Of the number of degrees of freedom represented by the number of past reports under consideration, some are absorbed by the selected time function representation; only the remainder are effective for smoothing. When it is known that the true positions necessarily follow a specified time function form, then the degrees of freedom devoted to this representation are exceedingly well spent, not only in regard to obtaining a good fit to existing points, but more important, in regard to the excellence of extrapolation.

On the other hand when there is no *a priori* knowledge of the functional form of the target trajectory, the time functions must be selected arbitrarily. The number of degrees of freedom that would be devoted to the functional representation would depend upon assumptions made regarding the inherent smoothness, or the spectral characteristics, of the motion of the target and accordingly would vary with the time interval covered by the past reports and the future projections. For time intervals over which relatively smooth behavior of the target trajectory cannot be expected, a relatively large number of degrees of freedom would have to be devoted to the functional representation. In any event, the number of degrees of freedom devoted to the functional representation would be, at least to some extent, a guess. In some cases the penalty for a bad guess could be severe and extrapolations may become rather tenuous. Hence in this case the advantage lies with the linear predictor which, as given in equation 1, does not restrict itself to a specified time function form, but instead allows such forms to develop in accordance with the statistical characteristics of the target's motion, and the errors of past position reports.

Because of the physical limitations on the target's maneuverability, the functional form of the variation of target co-ordinates with time can be specified with confidence when the time interval under consideration is short. Extension of accurate knowledge of the functional form to longer time periods would depend upon assumptions regarding the limitations which the target will choose to impose upon its maneuverability and cannot be projected with great confidence. A study of prediction over time intervals short enough to be treated with functional forms based upon physical limitations would probably be worth some effort. But this question will not be examined here. Present concern will be with predictions over time intervals long enough so that the target has the capability, and presumably, the probability of sufficient maneuvering so that its trajectory cannot be encompassed by simple functional forms. The properties of the predictor given in equation 1 will be examined with this purpose in mind.

It will be noted that in using equation 1 the problem at hand becomes analogous to that of the design of a linear noise filter.

THE NORMAL EQUATIONS

The selection of the coefficients, a_1, a_2, $\ldots$ a_n, to minimize the expected value of

$$(\mathbf{R}_0 - a_1\mathbf{R}_1 - a_2\mathbf{R}_2 \ldots a_n\mathbf{R}_n)(\mathbf{R}_0 - a_1\mathbf{R}_1 - a_2\mathbf{R}_2 \ldots a_n\mathbf{R}_n)$$

leads to the normal equations

$$\begin{pmatrix} \overline{\mathbf{R}_0\mathbf{R}_1} \\ \overline{\mathbf{R}_0\mathbf{R}_2} \\ \cdots \\ \overline{\mathbf{R}_0\mathbf{R}_n} \end{pmatrix} = \begin{pmatrix} \overline{\mathbf{R}_1\mathbf{R}_1} & \overline{\mathbf{R}_1\mathbf{R}_2} & \ldots & \overline{\mathbf{R}_1\mathbf{R}_n} \\ \overline{\mathbf{R}_1\mathbf{R}_2} & \overline{\mathbf{R}_2\mathbf{R}_2} & \ldots & \overline{\mathbf{R}_2\mathbf{R}_n} \\ \cdots & \cdots & & \cdots \\ \overline{\mathbf{R}_1\mathbf{R}_n} & \overline{\mathbf{R}_2\mathbf{R}_n} & \ldots & \overline{\mathbf{R}_n\mathbf{R}_n} \end{pmatrix} \times \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{pmatrix} \quad (2)$$

where the bar indicates average or expected values. These equations can be obtained by differentiating the average value of this expression with respect to each of the constants, $a_1, a_2, \ldots, a_n$. For a detailed description of the derivation of the normal equations see reference 1. Each target report can be expressed as

$$\mathbf{R}_i = \mathbf{r}_i + \boldsymbol{\epsilon}_i \quad (3)$$

where $\mathbf{r}_i$ is the true target position and $\boldsymbol{\epsilon}_i$ is the measurement error. Since $\mathbf{R}_0$ refers to the true future position for consistency of symbols we may write $\mathbf{R}_0 = \mathbf{r}_0$. It may be assumed that the measurement errors are uncorrelated with each other or the true target positions, so that

$$\overline{\boldsymbol{\epsilon}_i\mathbf{r}_j}=0$$

for all i, j and

$$\overline{\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_j}=0 \qquad (4)$$

for $i \pm j$.

Actually there would be some correlation in the magnitudes of $\boldsymbol{\epsilon}_i$ and $\mathbf{R}_i$ but if the measurement errors are unbiased there should be no correlation in direction, and this is sufficient for equation 4. The measurement errors may, and probably will be, directionally asymmetrical since errors in azimuth commonly are more significant than errors in range. Even so equation 4 follows when, as is to be hoped, the errors are unbiased. However, inclusion of knowledge of the departure of errors from directional symmetry might be worth considering in a more thorough analysis.

It will be assumed that $\overline{\boldsymbol{\epsilon}_i \cdot \boldsymbol{\epsilon}_i}$ is independent of i; that is to say, the rms measurement error is always the same. Then writing

$$\overline{\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i}=\sigma_e^2$$

and using equation 4, the normal equations may be written

$$\begin{pmatrix} \overline{\mathbf{r}_0\mathbf{r}_1} \\ \overline{\mathbf{r}_0\mathbf{r}_2} \\ \cdots \\ \cdots \\ \overline{\mathbf{r}_0\mathbf{r}_n} \end{pmatrix} = \begin{pmatrix} \overline{\mathbf{r}_1\mathbf{r}_1}+\sigma_e^2, & \overline{\mathbf{r}_1\mathbf{r}_2}, & \ldots & \overline{\mathbf{r}_1\mathbf{r}_n} \\ \overline{\mathbf{r}_1\mathbf{r}_2}, & \overline{\mathbf{r}_2\mathbf{r}_2}+\sigma_e^2, & & \\ \cdots & & & \\ \cdots & & & \\ \overline{\mathbf{r}_1\mathbf{r}_n}, & \overline{\mathbf{r}_2\mathbf{r}_n}, & & \overline{\mathbf{r}_n\mathbf{r}_n}+\sigma_e^2 \end{pmatrix} \times \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix} \qquad (5)$$

To define the normal equations completely it now remains only to evaluate $\overline{\mathbf{r}_i\mathbf{r}_j}$. Thus one is faced with the task of determining an autocorrelation function for target trajectories.

CONSIDERATIONS IN CALCULATING THE AUTOCORRELATION FUNCTION

The autocorrelation of target positions will depend upon the physical and stochastic properties of target maneuvers. If these are known, or assumed, the autocorrelation may be computed by numerical computation using Monte Carlo methods. However, to avoid the cost and effort required for such machine calculations, an attempt at analytic calculation will be made. As is to be expected for such an attempt, more or less drastic simplifying assumptions will be required.

Whatever method is used, it is important to note some additional factors that will affect the calculation. One such factor is the existence of knowledge regarding the target other than that contained in the past position reports. For example, if the target is identified as an aircraft of a given type, the specification of its maneuvering characteristics can be more restricted than when the identity of the target is not known. It would be manifestly difficult, if not impossible, to work out the autocorrelation functions for all possible, or expected targets. Also, application of such knowledge implies that one autocorrelation function will be used when the target is not identified and another will be employed after the target is identified, thus requiring a considerably more complicated predictor. However, there is one simple target characteristic that can be expected to have a significant effect on the autocorrelation function; specifically, its speed capabilities. This follows from the fact that almost any target that might be considered will be rather limited in its ability to change speed. Hence, a knowledge of the speed characteristics of the target, available from past observations, would play an important part in the calculation of the autocorrelation function.

Of course, some knowledge of the speed of the target is inherent in the existing target reports, but the problem of determining the speed of the target when only the existing reports are available is essentially different from that of determining the target's speed when a fairly accurate estimate is available from past observation. It is, therefore, generally envisioned that tracking procedures would be divided into two phases; the first, wherein a target is initially acquired and an initial estimate of speed is made and the second phase, which may be termed routine tracking, would be applied to targets when a good estimate of the speed already exists. It may be noted in passing that an estimate of target speed at the outset of a track may be very helpful in discarding "false alarms" identified by an unrealistic speed characteristic.

The attempt to calculate the autocorrelation function that will be made herein will be confined to the case of routine tracking. It will be assumed that the target speed is known. The problem will be further idealized by the assumption that the target does not change speed at all.

Another factor that influences the autocorrelation function is the statistics of the distribution in space of target births and deaths. To treat this matter in full detail would require considerable effort. First, there would be the treatment of the range limits of the radar sensing. This would be analogous, in the study of gaseous diffusion, to the effect of an interface between a liquid and its vapor, the interface absorbing molecules and also releasing them. Annihilation statistics would also play a part. Such statistics would be affected by the location of missile batteries and airfields and innumerable other factors.

Even if birth and death statistics could be handled in a reasonable manner, there would remain the problem of reconciling a predictor which always predicts a future position, without regard to life expectancy, with statistics embodying a finite life expectancy. In the present analysis no attempt will be made at treating life expectancy statistics in any detail. The question of life expectancy statistics will arise in connection with the problem of normalizing the observation space. This will be resolved in a somewhat arbitrary manner. However, since birth and death statistics of targets would also be somewhat arbitrary, it is hoped that assumptions made will produce no catastrophic errors.

NORMALIZATION

Consider first the problem of computing the value of $\overline{\mathbf{r}_i\mathbf{r}_1}$. Will this value change with t_i? If statistical properties are to be independent of time, that is to say, if the statistics of target distribution are stationary, then one would expect that this value should not change with time. On the other hand, if target life is taken to be indefinitely long, so that t_i is permitted to increase without limit, then the probability that $\mathbf{r}_i$ will be distant from the center of observations increases. Therefore, taking $\overline{\mathbf{r}_i\mathbf{r}_i}$ to be a finite constant places implicit restrictions on the probability of long target life. This is not of much consequence if the time interval under consideration is so short that the distance the target can move is quite small compared to the assumed constant value of $(\overline{\mathbf{r}_i\mathbf{r}_i})^{1/2}$, but it could lead to difficulties for longer time intervals. Therefore, it may be worthwhile not to restrict the value of $\overline{\mathbf{r}_i\,\mathbf{r}_i}$ to the true dimensions of the target observation space, but to introduce the fiction that the observation space is larger. It would be of interest to examine the case of an indefinitely large observation space. The results then can be expected to conform better with the intuitive conception that the predictor characteristics should be uniform over the observation space. If $\overline{\mathbf{r}_t\mathbf{r}_t}$ were taken to be finite or in particular to be in accord with the dimensions of the observation space, then boundary effects are implied which should be reflected in a difference in predictions for reports that differ only by a translation to or away from the origin.

Accordingly, it will be assumed that $\overline{\mathbf{r}_t\mathbf{r}_t}=r^2$ where r^2 is a constant, independent of t_t, and is eventually to be taken as arbitrarily large.

The assumed invariance of $\overline{\mathbf{r}_t\mathbf{r}_t}$ has some interesting implications. Thus it follows that

$$\overline{\mathbf{r}_j\mathbf{r}_j}-\overline{\mathbf{r}_t\mathbf{r}_t}=\overline{(\mathbf{r}_j+\mathbf{r}_t)(\mathbf{r}_j-\mathbf{r}_t)}$$
$$=\overline{\mathbf{r}_j(\mathbf{r}_j-\mathbf{r}_t)}+\overline{\mathbf{r}_t(\mathbf{r}_j-\mathbf{r}_t)}=0 \quad (6)$$

and hence

$$\overline{\mathbf{r}_j(\mathbf{r}_j-\mathbf{r}_t)}=\mathbf{r}_t\overline{-(\mathbf{r}_j-\mathbf{r}_t)} \quad (7)$$

At first glance this result may seem odd. Consider that $\mathbf{r}_j$ is the more recent target position, then it would seem that $(\mathbf{r}_j-\mathbf{r}_t)$ would be uncorrelated with $\mathbf{r}_t$. In other words, where a target goes would seem to be independent of where it has been. Hence it might be expected that the right-hand side of equation 7 would vanish. But surely both sides do not vanish for then,

$$\overline{(\mathbf{r}_j-\mathbf{r}_t)(\mathbf{r}_j-\mathbf{r}_t)}=\overline{\mathbf{r}_j(\mathbf{r}_j-\mathbf{r}_t)-\mathbf{r}_t(\mathbf{r}_j-\mathbf{r}_t)}$$

would vanish and this clearly is not true. Therefore, it is clear that $\overline{\mathbf{r}_t(\mathbf{r}_j-\mathbf{r}_t)}$ cannot vanish. This may be more readily accepted when it is expressed as follows: When a difference in target positions $(\mathbf{r}_j-\mathbf{r}_t)$ exists, it is more possible that $\mathbf{r}_t$ is in the direction $-(\mathbf{r}_j-\mathbf{r}_t)$ than in the direction of $+(\mathbf{r}_j-\mathbf{r}_t)$. With directions reversed this is also true for $\mathbf{r}_j$ and the assumption of constant $\overline{\mathbf{r}_t\mathbf{r}_t}$ leads to the equality, equation 7, between two nonvanishing quantities.

Since

$$\overline{(\mathbf{r}_j-\mathbf{r}_t)(\mathbf{r}_j-\mathbf{r}_t)}=\overline{\mathbf{r}_j\mathbf{r}_j}+\overline{\mathbf{r}_t\mathbf{r}_t}-2\overline{\mathbf{r}_t\mathbf{r}_j}$$
$$=2r^2-2\overline{\mathbf{r}_t\mathbf{r}_j} \quad (8)$$

the autocorrelation can be obtained from

$$\overline{\mathbf{r}_t\mathbf{r}_j}=r^2-1/2\overline{(\mathbf{r}_j-\mathbf{r}_t)(\mathbf{r}_j-\mathbf{r}_t)} \quad (9)$$

To determine $\overline{\mathbf{r}_t\mathbf{r}_j}$ it remains only to evaluate $\overline{(\mathbf{r}_j-\mathbf{r}_t)(\mathbf{r}_j-\mathbf{r}_t)}$

THE EVALUATION OF $\overline{(\mathbf{r}_j-\mathbf{r}_t\mathbf{r})\;({}_j-\mathbf{r}_t)}$

It may be assumed that the value of $\overline{(\mathbf{r}_j-\mathbf{r}_t)(\mathbf{r}_j-\mathbf{r}_t)}$ is not dependent upon absolute time but is a function only of the time difference $\tau_{jt}=|t_j-t_t|$. The value of this quantity is assumed to be independent of $\mathbf{r}_t$ so that in this calculation, for convenience, one may take $\mathbf{r}_t=0$. The value sought then is the mean-square-distance, $\sigma^2(\tau)$, travelled by a target initially at $\mathbf{r}=0$ in the time $\tau=\tau_{jt}$. Since the direction of the initial velocity is immaterial, it may be chosen arbitrarily, for definiteness, say in the direction of the x-axis. Let θ be the angle between the direction of target's velocity and the x-axis.

Then initially $\theta=0$, and

$$\sigma^2(\tau)=\overline{x^2(\tau)}+\overline{y^2(\tau)}$$
$$=\overline{V^2\int_0^\tau \cos\theta(s)ds\int_0^\tau \cos\theta(r)dr}+$$
$$\overline{V^2\int_0^\tau \sin\theta(s)ds\int_0^\tau \sin\theta(r)dr}$$
$$=V^2\int_0^\tau\int_0^\tau \overline{\cos\,[\theta(s)-\theta(r)]}ds\,dr \quad (10)$$

where on the right-hand side the order of integration and averaging has been interchanged.

Because stationary statistical characteristics are assumed it follows again that $\overline{\cos\,[\theta(s)-\theta(r)]}$ is independent of the absolute value of r and s, but dependent on the value of the time difference $T=|s-r|$ and is assumed independent of the absolute value of $\theta(r)$. The latter assumption implies that the probability of turning is independent of time and past history which may not be strictly true, but is used here as a workable approximation. To assume otherwise would mean that in the tracking process attention should be given to the value of $\dot\theta$ in addition to V, and that some knowledge is available regarding the persistence of $\dot\theta$. The view taken here, that θ might be changed at any time, corresponds to ignoring limitations on $\dot\theta$.

In accordance with the foregoing assumptions, $\overline{\cos\,[\theta(s)-\theta(r)]}$ is taken to be the value of $\overline{\cos\theta(T)}$, for targets for which $\theta(0)=0$. The value of $\overline{\cos\theta(T)}$ can be computed if the probability distribution of $\theta(T)$ is known. To obtain an estimate of this probability distribution, consider the variations of the angle θ in time to be analogous to the one-dimensional random walk problem. Accordingly, the probability density $p[\theta(T)]$ would have the form[2,3]

$$p[\theta(T)]=\frac{1}{(2\pi T)^{1/2}}e^{-\frac{\theta^2}{2\alpha^2T}} \quad (11)$$

In the treatment of gaseous diffusion by the random-walk analogy, it is found that the quantity α^2 is approximately given by the product of the mean free path, the mean speed and a proportionality constant that is close to unity. Analogously, in the case of equation 11, α^2 would be expected to be given approximately by the product of the mean θ excursions between $\dot\theta$ changes, say $\overline{|\theta|}$, and the mean value of $\overline{|\dot\theta|}$. Thus

$$\alpha^2=\overline{|\theta|}\,\overline{|\dot\theta|} \quad (12)$$

As is characteristic of normal distributions, equation 11 cannot be true literally, since it extends to infinite values of θ in finite time, but if the value of α is taken in recognition of the physical limitations on the target's maneuvering capabilities, then the probability of impossibly large values of θ will be negligible. The physical limitations on $\overline{|\theta|}$ are fairly clear. Since only a finite force is available for turning (or as is said, the target can "pull" only a limited number of g's in a turn, generally 3 or less) $\overline{|\dot\theta|}$ is bounded, and in general

$$|\dot\theta|<\frac{3g}{V} \quad (13)$$

The boundness of $\overline{|\theta|}$ follows from the limitations on $\dot\theta$ and the implicitly finite target life.

It would appear, however, that the value of α, while restricted by physical limitations, would be more directly related to the choice the target makes in regard to maneuvering and evasion tactics. Its purpose in life is not to execute evasion maneuvers but to achieve an objective that becomes more difficult if not impossible if it evades excessively. This means that while it has the capacity to make α fairly large by taking headings at large angles to the direction to the objective, it will accept a smaller value of α so as not to prejudice its ability to accomplish its mission. It appears reasonable to assume that it is unlikely to hold headings much more than 10 degrees removed from its average course, so that one may write

$$\overline{|\theta|}<10° \quad (14)$$

Equations 13 and 14 serve to put a bound on α. Rather than assign a specific value to α for study purposes, one may choose to vary α using it as a parameter to describe the extent to which maneuvering and evasion tactics are expected of targets. In this way the factors in the tracking process that influence ability to follow maneuvers may be assessed. For such a study, it is convenient to have a bound on α so that the study is not extended to unrealistically large values.

Assuming the value of α to be known, or suitably chosen, $\overline{\cos\theta(T)}$ may be computed using equation 11. Thus,

$$\overline{\cos\theta(T)}=\frac{1}{(2\pi T\alpha)^{1/2}}\int_{-\infty}^{\infty}\cos\theta e^{-\frac{\theta^2}{2\alpha^2T}}d\theta$$

$$=e^{-\frac{\alpha^2T}{2}} \tag{15}$$

Hence

$$\overline{\cos\,[\theta(s)-\theta(r)]}=e^{-\frac{1}{2}\alpha^2|s-r|} \tag{16}$$

and from equation 10

$$\sigma^2(\tau)=V^2\int_0^\tau\int_0^\tau e^{-\frac{1}{2}\alpha^2|s-r|}dr\,ds$$

$$=2V^2\int_0^\tau\int_0^s e^{-\frac{1}{2}\alpha^2|s-r|}dr\,ds \tag{17}$$

$$=\frac{4V^2}{\alpha^2}\int_0^\tau e^{-\frac{1}{2}\alpha^2s}\left(e^{\frac{1}{2}\alpha^2s}-1\right)ds$$

$$=\frac{8V^2}{\alpha^4}\left(\frac{\alpha^2\tau}{2}-1+e^{-\frac{\alpha^2}{2}\tau}\right)$$

Hence

$$\overline{(\mathbf{r}_j-\mathbf{r}_i)(\mathbf{r}_j-\mathbf{r}_i)}=\sigma^2(\tau_{ij})=\frac{8V^2}{\alpha^4}f\left(\frac{\alpha^2}{2}\tau_{ij}\right) \tag{18}$$

where $f(x)=(x-1+e^{-x})$ and as previously mentioned $\tau_{ij}=|\tau_i-\tau_j|$.

As would be expected, the function $f(x)$ is monotonically increasing. Its first and second derivatives are always positive for $x>0$. When $x\to0$, $f(x)\to x^2/2$ so that

$$\overline{(\mathbf{r}_j-\mathbf{r}_i)(\mathbf{r}_j-\mathbf{r}_i)}=V^2\tau_{ij}{}^2 \tag{19}$$

This corresponds to the case of a nonmaneuvering target, where $\alpha\to0$. The same result applies when the time interval covered is extremely short in which case the target does not have time to maneuver.

Calculation of the Coefficients and the Standard Error of Estimate

The completion of the solution is now more or less straightforward. Using equations 9 and 18, the autocorrelation is given by

$$\overline{\mathbf{r}_i\mathbf{r}_j}=r^2-\frac{4V^2}{\alpha^4}f\left(\frac{\alpha^2\tau_{ij}}{2}\right) \tag{20}$$

This result may be used in the normal equation 5 and the values of the coefficients a_1, a_2, … a_n, found for cases of interest. That the values of these coefficients will converge as r^2 grows indefinitely large may be seen as follows:

The solution of the normal equations can be expressed as the ratio of two determinants using Cramer's rule. All elements of both determinants are of the form: r^2+c_{ij} where $c_{ij}=\delta_{ij}\sigma^2-1/2\sigma^2\tau_{ij}$ (and δ_{ij} is Kronecker's delta). The determinants can be expanded as a sum of determinants having rows (or columns) whose elements are either c_{ij} or are all equal to r^2. Those determinants having two or more rows (or columns) of r^2 will vanish, hence only determinants having one or no rows (or columns) of r^2 will contribute and the solution for the coefficients can be put in the form

$$a_i=\frac{\ell_ir^2+m_i}{n_ir^2+\sigma_i} \tag{21}$$

Where ℓ_i, m_i, n_i and σ_i are independent of r^2 and in general finite. As r^2 becomes large the coefficients approach the values

$$a_i\to\frac{\ell_i}{n_i}\left[1+\left(\frac{m_i}{\ell_i}-\frac{\sigma_i}{n_i}\right)\frac{1}{r^2}\right]$$

$$=p_i+q_i\frac{1}{r^2}\to p_i \tag{22}$$

where

$$p_i=\frac{\ell_i}{n_i}$$

and

$$q_i=\frac{\ell_i}{n_i}\left(\frac{m_i}{\ell_i}-\frac{\sigma_i}{n_i}\right)$$

For present purposes the values $a_i=p_i$ may be used for the linear predictor.

Once the coefficients are known the standard error of estimate say, δ, maybe computed as follows:

$$\delta^2=\overline{(\mathbf{R}_0-a_1\mathbf{R}_1-a_2\mathbf{R}_2-\ldots-a_n\mathbf{R}_n)(\mathbf{R}_0-a_1\mathbf{R}_1-a_2\mathbf{R}_2-\ldots-a_n\mathbf{R}_n)} \tag{23}$$

and noting that the coefficients satisfy equation 2.

$$\delta^2=\overline{\mathbf{R}_0\mathbf{R}_0}-a_1\overline{\mathbf{R}_0\mathbf{R}_1}-a_2\overline{\mathbf{R}_0\mathbf{R}_2}-\ldots-a_n\overline{\mathbf{R}_0\mathbf{R}_n}$$

$$=\overline{\mathbf{R}_0\mathbf{R}_0}-\sum_{i=1}^n a_i\overline{\mathbf{R}_0\mathbf{R}_i} \tag{24}$$

whereupon using equations 3, 4, 9, and 18

$$\delta^2=r^2+\sigma_e{}^2-\sum_{i=1}^n a_i\overline{\mathbf{r}_0\mathbf{r}_i}$$

$$\delta^2=\left(1-\sum_{i=1}^n a_i\right)r^2+\sigma_e{}^2+$$

$$\frac{4V^2}{\alpha^4}\sum_{i=1}^n a_i\left(\frac{\alpha^2\tau_{0i}}{2}-1+e^{-\frac{\alpha^2}{2}\tau_{0i}}\right) \tag{25}$$

Since the coefficients minimize δ^2 under the condition that r^2 is indefinitely large it is clear that the a_i must converge to values such that

$$\sum_{i=1}^n a_i=\sum_{i=1}^n p_i=1 \tag{26}$$

This relationship confirms the uniformity of the predictor in regard to the observation space, since it implies that the predicted position translates with any translation of the past position reports. (Insensitivity to the rotation of the past position reports follows from the use of the same linear predictor for the x and y co-ordinates.)

Using equations 22 and 26 the expression for δ^2 becomes, for indefinitely large r^2

$$\delta^2=\sum_{i=1}^n q_i+\sigma_e{}^2+\frac{4V^2}{\alpha^4}\sum_{i=1}^n a_i\times$$

$$\left(\frac{\alpha^2\tau_{0i}}{2}+e^{-\frac{\alpha^2}{2}\tau_{0i}}\right) \tag{27}$$

APPLICATIONS

The foregoing analysis can be applied to general exploratory studies as well as to cases of special interest. A general study could be executed most readily for the case where the reports and the prediction are evenly spaced in time. Then the parameters that enter the problem are:

Δt = the time interval between reports
α = the maneuverability parameter
V = the target velocity
σ_e = the rms measurement error
n = the number of past position reports that are retained

The coefficients of the predictor are nondimensional. The only other significant result, the standard error of estimate can be put in nondimensional form by expressing it as a fraction of σ_e, $V\Delta t$, or V/α^2. The nondimensional results can only be functions of the nondimensional parameters that can be formed from those just listed. Only three independent parameters can be formed: These may be chosen in a variety of ways. For example, the choice might be:

$$\frac{\alpha^2\Delta t}{2},\ \frac{4V^2}{\alpha^4\sigma_e},\ n$$

Where the predicted position does not conform to the even time spacing then an additional nondimensional parameter will be required. This may be chosen to be the ratio of the projection of time into the future to the time interval between past reports.

When only three or four nondimensional parameters are involved a rather complete study would appear to be feasible. For the case of arbitrarily spaced reports the number of parameters becomes excessive if more than a few past reports are considered. Also the implimentation of a predictor incorporating the results for this situation would be more difficult. However, the calculation of some representative cases may be effective in demonstrating how a proposed predictor departs from what could be obtained, ideally, and for guiding the design of a suitable predictor.

The selection of gate sizes would be predicated upon the value expected for the standard error of estimate, and upon the probability of "escape" that is deemed to be tolerable. An estimate of the probability of escape would be computed upon the basis of an assumed normal probability distribution for the deviation from the prediction. However, it should be kept in mind that such calculations of escape probability do not include the effect of departures of target's motion from the idealized characteristics. The most important factors in this connection are the assumed two-dimensional motion and the constancy of V. Minor variations in V and moderate changes in altitude should work no hardship, and probably, if desired, modifications could be introduced to encompass these. However the possibility remains of drastic departures from the idealized characteristics. Thus when an airplane executes an acrobatic maneuver such as a loop, its track would probably be interrupted. Likewise, a tracking system such as described herein could not be expected to cope with the maneuvers of a heliocopter.

In designing a tracking system, it must be recognized that the probability of escape will not be reduced to zero; at least not with any simple mechanization. A reasonable approach is first to seek to reduce escape probabilities to small values as simply and conveniently as possible. The basic portion of the tracking system should therefore be designed to accommodate the usual or most commonly expected target characteristics. If proceeding in this way the number of fugitives can be kept small, then it may be feasible to institute more elaborate search methods to recapture lost targets should this be needed and/or desired.

References

1. Numerical Mathematical Analysis, James B. Scarborough. John Hopkins Press, Baltimore, Md., 1950.

2. Kinetic Theory of Gases, Earle H. Kennard. McGraw-Hill Book Company, Inc., New York, N. Y., 1938.

3. An Introduction to Kinetic Theory of Gases, Sir James Jeans. Cambridge University Press, Cambridge, England, 1952.

Global Public Telephone Service—1958

D. D. DONALD
MEMBER AIEE

T. A. CHANDLER
NONMEMBER AIEE

AT the Symposium on Global Communications in Washington, D. C., in June of 1954, Mr. Donald of the American Telephone and Telegraph Company presented a paper discussing the global public telephone service as provided by that company. It is proposed here to review some aspects of public telephone service in the light of the important developments that have taken place since then.

Overseas Message Telephone Growth

Overseas public telephone service between the United States and other areas in the world has continued to grow since 1953, not only in the number of areas reached, which has increased more than 20%, but also in the number of overseas public telephone messages which are handled. The volume of messages has increased approximately 87% over the number handled in 1953 as can be seen from Fig. 1 which shows the total overseas public messages between the United States mainland and other areas by years. It will be noted that the placing in service of the transatlantic telephone cable system between the United States and Great Britain in September 1956 contributed very substantially in the growth of traffic experienced. As a matter of interest, one year after the transatlantic cable system was in service the volume of messages with Great Britain had increased approximately 90%, with substantial increases also being experienced with those countries on the European continent which are served by direct circuits in this cable system. Another point of interest is that after the cable system had been in operation, it was noted that the average length of conversations between the United States and Great Britain had increased approximately 20% and a similar but somewhat smaller increase was experienced between the United States and countries in Europe which are served by the cable.

The first telephone cable system across the Atlantic Ocean was followed by a similar system between Seattle, Wash., and Ketchikan, Alaska, which was placed in service in December 1956. Again, a marked increase in the volume of traffic occurred with the result that one year later it had increased approximately 30%. A third underseas telephone cable system was placed in service in October 1957 between the mainland of the United States and Hawaii. As in the two previous cable systems, this sytem also shows a substantial increase in the volume of telephone traffic handled. A comparison of the traffic volume for several months prior to the introduction of service over this cable system and the same period one year later shows an increase in the order of 50%.

Fig. 2 shows the various overseas points which are connected with the Bell System to furnish public telephone service. The dots indicate points served by direct circuits or links and the circles are points served by switched circuits; for example, service with the Canary Islands is furnished via Madrid. It can be seen that there are relatively few areas in the world with which service is not available.

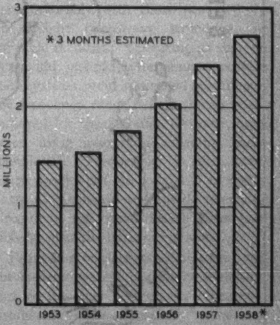


Fig. 1. Overseas messages

Paper 59-920, recommended by the AIEE Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted December 26, 1958; made available for printing May 11, 1959.

Douglas D. Donald and Thurlow A. Chandler are with The American Telephone and Telegraph Company, New York, N. Y.

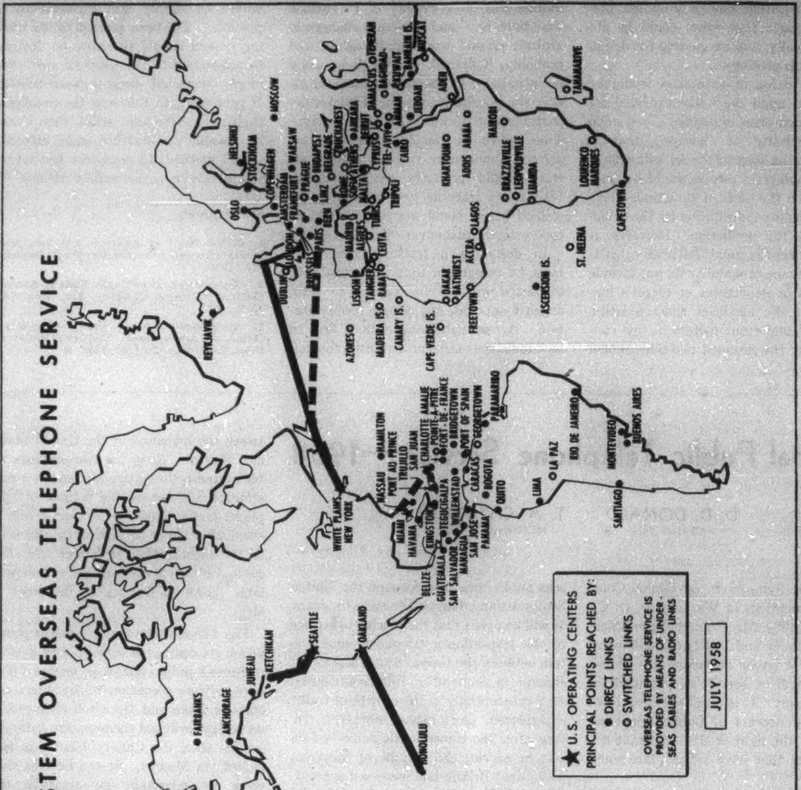

BELL SYSTEM OVERSEAS TELEPHONE SERVICE
★ U.S. OPERATING CENTERS
PRINCIPAL POINTS REACHED BY:
● DIRECT LINKS
○ SWITCHED LINKS
OVERSEAS TELEPHONE SERVICE IS PROVIDED BY MEANS OF UNDER-SEAS CABLES AND RADIO LINKS
JULY 1958
REYKJAVIK
FAIRBANKS
ANCHORAGE
JUNEAU
KETCHIKAN
SEATTLE
OAKLAND
HONOLULU
ADAK
WHITE PLAINS
NEW YORK
MIAMI
HAVANA
NASSAU
PORT AU PRINCE
TRUJILLO
SAN JUAN
HAMILTON
CHARLOTTE AMALIE
POINTE-A-PITRE
FORT-DE-FRANCE
BRIDGETOWN
PORT OF SPAIN
GEORGETOWN
PARAMARIBO
KINGSTON
TEGUCIGALPA
WILLEMSTAD
CARACAS
BOGOTA
QUITO
SAN JOSE
PANAMA
MANAGUA
SAN SALVADOR
GUATEMALA
BELIZE
LIMA
LA PAZ
RIO DE JANEIRO
MONTEVIDEO
BUENOS AIRES
SANTIAGO
AZORES
MADEIRA IS.
CANARY IS.
CAPE VERDE IS.
DAKAR
BATHURST
FREETOWN
ACCRA
LAGOS
ASCENSION IS.
ST. HELENA
CAPETOWN
LOURENCO MARQUES
BRAZZAVILLE
LEOPOLDVILLE
LUANDA
KHARTOUM
ADDIS ABABA
NAIROBI
TANANARIVE
ADEN
JEDDAH
DHAHRAN
BAHRAIN IS.
KUWAIT
BAGHDAD
TEHERAN
DAMASCUS
AMMAN
BEIRUT
TEL-AVIV
CYPRUS
CAIRO
TRIPOLI
TUNIS
MALTA
ANKARA
ISTANBUL
ATHENS
SOFIA
BUCHAREST
BELGRADE
BUDAPEST
ROME
LINZ
BERN
PARIS
PRAGUE
FRANKFURT
WARSAW
AMSTERDAM
COPENHAGEN
STOCKHOLM
HELSINKI
OSLO
MOSCOW
LONDON
DUBLIN
BRUSSELS
MADRID
LISBON
ALGIERS
TANGIER
RABAT
CEUTA
KABUL
KARACHI
BOMBAY
GOA
COLOMBO
BANGKOK
HONG KONG
MACAO
SAIGON
SINGAPORE
BANDUNG
SHANGHAI
TAIPEI
MANILA
SEOUL
OKINAWA
TOKYO
GUAM
HOLLANDIA
RABAUL
PORT MORESBY
SYDNEY
HOBART
NEW CALEDONIA
SUVA
WELLINGTON

Fig. 2. Overseas points of communication and cable routes

Underseas Public Telephone Cable Systems

The underseas telephone cable systems between the United States mainland and Great Britain, Alaska, and Hawaii respectively, are indicated by the solid lines in Fig. 2. The dashed lines on this figure show systems planned to the European continent and to Puerto Rico. The basic design of these systems is the same in that a pair of cables is used, one in each direction of transmission, each is designed to provide 36 voice channels of normal bandwidth, uses flexible repeaters built into the cable, and is powered from the cable terminal stations.

The transatlantic cable system terminals are at White Plains, N.Y., Montreal, Que., Canada, and London, England, with the terminals of the deep-sea cable portion located only at Clarenville, Newfoundland, and Oban, Scotland. This system is a joint project of the General Post Office in England, the Canadian Overseas Telecommunication Corporation, and the American Telephone and Telegraph Company. Of the 36 voice channels in the system $29^1/_2$ are allocated for United States traffic and $6^1/_2$ for Canadian traffic.

From the deep-sea cable terminal at Clarenville, the United States channels are extended to White Plains, by means of a submarine cable system between Clarenville and Sydney Mines, Nova Scotia; by a *TD-2* microwave-radio relay system from the latter point to Portland, Maine; and by cable-carrier groups to White Plains. The Canadian channels branch off at Spruce Lake, a station on the *TD-2* route in Canada, and are extended to Montreal on facilities comparable to those in the United States. In the United States group, a number of the circuits are extended from London to points on the European continent for service with those countries. As the channels in this system are loaded to capacity, discussions are under way between the partners in this project toward increasing the number of voice channels to handle the continuing growth in traffic.

The underseas portion of the Alaska system extends from Port Angeles, Wash., to Ketchikan, Alaska, with an extension at the United States end to Seattle, where it connects with the domestic long-distance network. At Ketchikan, the 36 voice channels connect to facilities of the Alaska Communications System to furnish telephone service to various points throughout Alaska. Between Ketchikan and Skagway in Alaska, service is furnished over Alaska Communications System facilities consisting of a single underseas cable using land-based repeaters.

The underseas telephone cable system with Hawaii is a joint project of the American Telephone and Telegraph Company and the Hawaiian Telephone Company. The deep-sea cable terminals are located at Point Arena, Calif., and Hanauma Bay on the island of Oahu, with the system terminals at Oakland and Sacramento, Calif., and Honolulu. The deep-sea portion of this system is the longest of the three in operation, extending approximately 2,540 miles with 57 cable repeaters in each direction. With the commencement of service over this cable system, overseas dialing was inaugurated simultaneously. On the United States mainland any long-distance operator with access to the dialing network can dial a call to any number in Hawaii connected to the dialing network in that area and similarly, in the opposite direction from Hawaii to the United States mainland. As the channels in this cable are equipped for dialing and a number of the radio circuits between Oakland and Honolulu are also maintained in operation, it was necessary to equip the radio circuits for dial operation as is discussed later in this paper.

These cable systems provide high-quality telephone circuits which meet the same transmission requirements as domestic circuits of comparable lengths. These circuits are free from fading and have a low noise level which is not subject to fluctuations due to radio propagation variations. The transmission-frequency characteristics are stable as contrasted with the variations encountered in high-frequency radio circuits resulting from selective fading. They are free from the usual propagation outages on radio that result from magnetic storms and sudden ionospheric disturbances particularly on the North Atlantic and Alaska routes.

With the success of the three undersea telephone cable systems already in operation it became evident that additional cable systems would be required wherever they were needed to handle large volumes of traffic and could be justified from an economic viewpoint. Plans have been made and construction has commenced on a second underseas telephone cable system across the Atlantic Ocean between the United States and the continent of Europe which is planned for service in late 1959. Partners in this project are the Government Administrations of France and Germany, together with the American Telephone and Telegraph Company. The deep-sea cable terminals for this system will be at Clarenville, Newfoundland, and Penmarch, France. Another system is planned between the United States mainland and Puerto Rico for service in early 1960. This will be a joint project of the American Telephone and Telegraph Company and the Radio Corporation of Puerto Rico, a subsidiary of the International Telephone and Telegraph Corporation. The cable terminals of this system will be at West Palm Beach, Fla., and San Juan, Puerto Rico. The basic design of these two systems will be the same as the three already in service.

As satisfactory results were obtained with traffic operator dialing on the circuits in the Hawaiian cable system, plans have been made to introduce dialing on the Alaska circuits early in 1959. It is planned to introduce dialing on the Puerto Rico circuits when that underseas cable system is placed in service.

Over-the-Horizon Radio

The development of over-the-horizon or tropospheric-scatter radio techniques for multichannel telephone systems has been a significant development in the last few years. This technique permits the establishment of radio links of 200 miles or longer depending on the number of voice channels desired, using radio frequencies in the uhf (ultrahigh-frequency) range. Such systems, when carefully engineered, are capable of providing highly reliable service for a group of telephone circuits up to about 200 in number. Relatively high-power transmitters and large antennas are required for the longer distances and larger numbers of circuits.

The great advantages of this technique lie in the high reliability of the circuits as compared to circuits using high-frequency radio techniques, especially in regions of the world subject to great ionospheric propagation instability. The quality of the circuits provided compares very favorably with those provided by line-of-sight microwave techniques or by cable facilities. Operating as they do in the uhf part of the frequency spectrum such systems avoid additional crowding of the frequency range 2–7 mc (megacycles) required for the 200-mile distance range when using ionospheric-reflection radio techniques. Frequencies in this latter range would not be available in many parts of the world to provide telephone circuits in the quantities that are possible by over-the-horizon systems.

From an economic standpoint the over-the-horizon system is usually somewhat less costly than a deep-sea repeatered submarine cable especially when the number of telephone circuits approach the maximum which the system is ca-

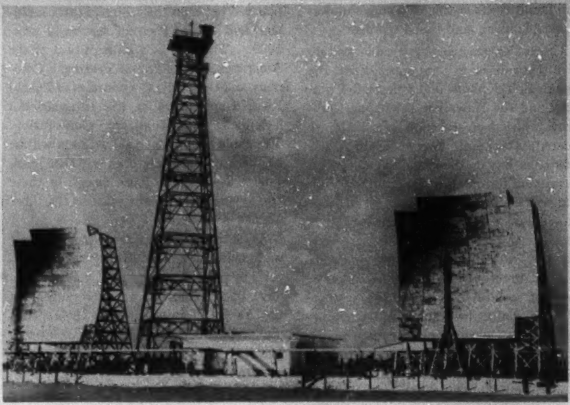


Fig. 3. Florida City Station with 60-foot antennas and microwave tower

pable of handling. For small numbers of circuits, radio using the ionospheric-reflection techniques is far less expensive but does not approach over-the-horizon techniques in quality and reliability. Where line-of-sight radio systems using intermediate repeaters are readily installed and maintained, it appears that such systems will usually be less costly.

An important radio link of this type has been provided between Florida City, Fla., and Guanabo, Cuba, by the American Telephone and Telegraph Company on the Florida end and the Radio Corporation of Cuba, a subsidiary of the International Telephone and Telegraph Corporation, on the Cuban end. This system has been described at other meetings and it is proposed here to only note its general features.

This system employs 10-kw transmitters and 60-foot parabolic reflector antennas. Two channels are provided for full protection of service. Each channel employs space diversity reception and diversity combining of the two channels is also provided. At the United States end a 2-link *TD-2* microwave system connects to Miami and at the Cuban end a similar single link connects to Havana. Fig. 3 shows a general view of the Florida City Station indicating the 60-foot parabolas and the *TD-2* tower. Fig. 4 shows the route layout of the system. This figure also shows the proposed system to Nassau, Bahamas, which is planned for service by the end of 1959. Whereas the Cuban system operates in the 700–900-mc band, it is proposed to operate the Nassau system in the 2,000–2,200-mc band. Thirty-foot parabolic antennas are being employed in the latter system.

The Cuban system at present is providing 36 telephone circuits of the same general grade of quality and reliability as are provided by the cable circuits on the route to Cuba. It is capable of an ultimate capacity in excess of 100 circuits when growth requires this number. Video service is also being provided over this system on an occasional basis, as required, making use of the second radio channel which is provided for protection. Because of selective fading, video service is not entirely comparable to that provided over line-of-sight microwave systems, but monochrome signals acceptable for broadcasting are being received over the system with a high degree of reliability.

Other systems employing over-the-horizon techniques have been provided elsewhere in the world. The most extensive of these used for public telephone service is the Alaska Communication System project called White Alice. This is a joint military and public telephone service system. Other systems used for public telephone service have been installed in Canada, the Caribbean area, and in the European area.

High-Frequency Radio

High-frequency radio continues to provide the facilities for most of the transoceanic links in the world telephone network. Although the use of cables on major traffic routes and over-the-horizon links, in a few cases where they are feasible, will continue to increase, it appears that high-frequency ionospheric-reflection techniques will continue to play a major role in the international public telephone service for many years to come. They are necessary to serve many places where cable is not feasible and over-the-horizon radio techniques are not applicable. They are required to parallel cable routes to provide protection and dispersion of facilities. These radio facilities also provide direct connection between countries where the cable routes traverse intermediate countries. It is urgently desirable that these facilities be improved in quality and reliability. A number of trends are evident in this direction.

The use of single-sideband techniques continues to expand and provides an improvement in service both in signal-to-noise ratio for equivalent power output and a quality improvement. The most important advantage of this technique

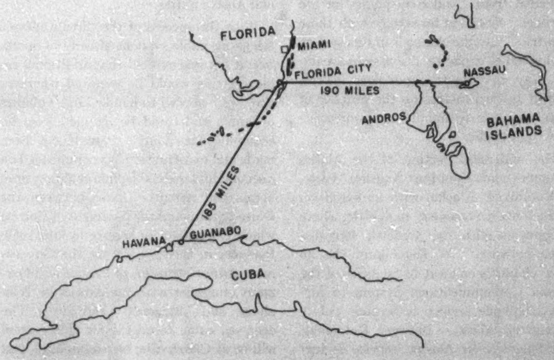


Fig. 4. Over-the-horizon system routes at Florida City

is, however, the conservation in spectrum occupancy which it provides. The overseas public radiotelephone circuits terminating in the United States are approximately 95% on a single-sideband basis now and this percentage is increasing steadily.

There is also a trend toward the use of higher power on transoceanic circuits. In this connection the American Telephone and Telegraph Company has recently placed in service ten newly designed 30-kw power amplifiers. The provision of additional amplifiers is planned for services where a significant improvement in service continuity can be obtained.

This is a single-stage amplifier designed by the Westinghouse Electric Corporation to Bell System specifications. It operates in the frequency range 4 to 26.5 mc, utilizing two Machlett type *6423* air-cooled vacuum tubes connected in parallel and operated in a grounded grid circuit. A peak envelope power of 30 kw is provided into a balanced load of from 200 to 900 ohms resistance when driven by an exciter capable of an output of 2.5-kw peak envelope power. The amplifier output pi-network is somewhat unique in that there are no roller or sliding contacts, the tuning being accomplished by means of fixed coils and variable vacuum capacity, the required combination being switched in by means of vacuum relays.

A servosystem is incorporated in the amplifier and is arranged for interconnection with the servosystem of the Western Electric Company *LD-T2* transmitter used as a driver in such a way that any one of ten preset frequencies may be selected by pushbutton control either at the equipment or, with suitable circuits, from a remote point. Fig. 5 gives a view of part of the transmitter room at the Lawrenceville, New Jersey, station showing these amplifiers with the *LD-T2* driver units.

Fig. 5. Amplifiers at the Lawrenceville, N. J., transmitting station

These are being used on the European paths which approach the Auroral Zone and are most in need of improvement. Most of the European countries where these circuits terminate have also installed equipment of comparable power output.

High-Frequency Radio Spectrum

With the increasing number of fixed radio services operating in the high-frequency range it has been necessary to take advantage of all available techniques to obtain improved utilization of the spectrum and conserve the available frequency space. Typical of such measures is the extensive use which has been made for many years of the single-sideband method of operation and the use of multiplex systems for both telephony and telegraphy.

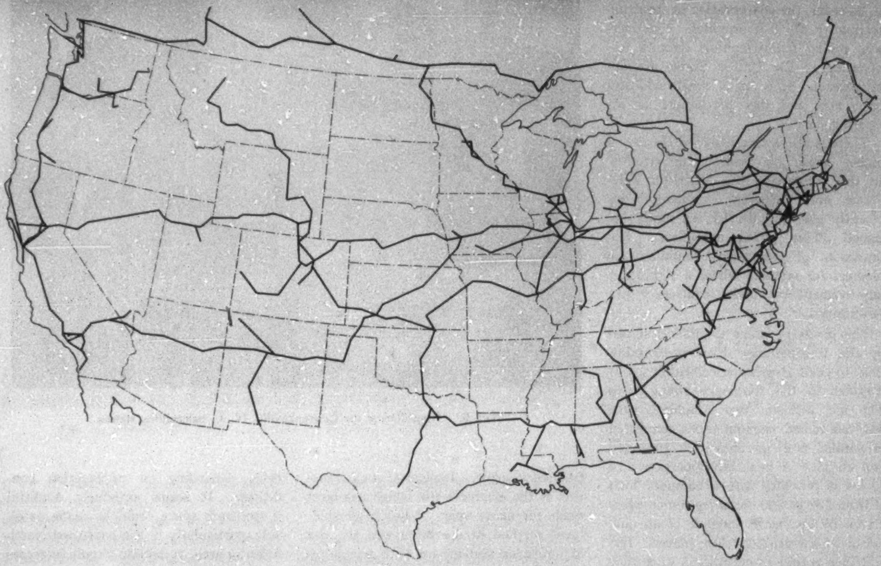
In the paper referred to in the opening paragraph, the results of an analysis showing the number of radio-frequency assignments in the ranges 7,300 to 8,195 kc and 14,350 to 14,990 kc were given. A recent analysis of the same frequency ranges has been made and it shows that there had been an increase of approximately 35% in the number of stations which have registered for operation in these bands. On the average, there are more than 13 stations registered for every kilocycle in the 7,300–8,195-kc range and more than two stations for every kilocycle in the 14,350–14,990-kc range. Considering that each telephone circuit requires a minimum of 3 kc of spectrum space, it can be seen that the likelihood of finding additional frequencies which are clear of interference to provide service expansion is extremely difficult in most parts of the world.

As the overseas public telephone service is continuing to grow, the provision of additional telephone circuits by high-frequency radio using the most efficient multichannel single-sideband techniques would require the continuous use of 108 kc of spectrum space in each direction of transmission to furnish 36 circuits, or the basic capacity of one underseas cable system. This would, depending on the route, require assignments totaling 216 kc in each of four or five different frequency ranges, each being used part-time, depending on propagation conditions. It seems extremely doubtful if spectrum space could be made available, particularly in the congested North Atlantic area, to handle circuit increases of this magnitude.

Line-of-Sight Microwave Radio-Relay Systems

Another significant development in recent years has been the rapid expansion in the use of line-of-sight microwave radio-relay systems in the public telephone networks. This has been particularly true in the United States and Canada and there is considerable use of these systems in Europe and other parts of the world.

Fig. 6 indicates the major microwave radio-relay routes of the Bell System and connecting carriers used for public telephone service in the United States and Canada. It will be noted that the systems extend from coast-to-coast and from Canada to Southern United States. Most of these routes employ the Western Electric *TD-2* radio system. Many sections have been developed to the full capacity of six channels in each direction for telephone and television purposes. The individual radio channels used for telephony are generally developed for about 600 telephone circuits. In deriving the telephone circuits a frequency-division single-sideband carrier technique is used employing the same carrier systems as are used on coaxial cables. The combined carrier signal is transmitted over the radio system by frequency modulation.



OCTOBER 1, 1958

Fig. 6. Broad-band microwave radio-relay system

Vol.
Pt. 1
1959

In the United States a large part of the new long-distance telephone facilities provided in recent years have employed microwave facilities. Experience indicates that this is a highly reliable and economical method of handling telephone service as well as the network television service.

Overseas Dialing

The growth in traffic in the overseas service has made it desirable to consider improved methods of handling calls. Experience in the United States and other countries indicates that calls can be more speedily and accurately established by properly engineered dialing systems. Rapid strides have been made in the United States in applying dial signaling systems to the long-distance network. It was only natural that as the business to Hawaii grew that consideration should be given to extending this method to that part of the United States.

The transmission of dial signaling information and supervisory signals over the underseas telephone cables presents no new problems. The same general methods as are used throughout the United States long-distance network can be employed with appropriate terminal facilities to integrate them with the local dial facilities.

The transmission of signaling information with adequate reliability is a problem on high-frequency radio circuits. It is well known that on high-frequency radio circuits using ionospheric reflections, any single frequency signal is subject to more or less continuous variations in received level due to multipath fading. In telegraph systems these are largely overcome by diversity and frequency shift methods.

The requirements of a dial signaling system for high-frequency radiotelephone circuits are that the signaling tones must be contained within the normal bandwidth of the telephone circuit to conserve the use of the frequency spectrum, the tones must not be transmitted continuously so that the radio transmitter will not be loaded excessively, and the required signals must be received with adequate reliability during all radio propagation conditions during which the telephone circuit is suitable for conversation by the telephone-using public.

A signaling system meeting these requirements has been developed and is being used successfully on the radiotelephone circuits to Hawaii. This sends tones at 800 and 1,900 cycles which are within the telephone band during intervals when signals are required. Tones are sent in duplicate, spaced 1,100 cycles apart to provide frequency diversity, and dial pulses are sent by frequency shift where the frequency of both tones are shifted 200 cycles upward in frequency.

This method is applicable to any high-frequency radiotelephone route where dial signaling is desired. There are a number of problems in introducing dial signaling to foreign countries. When these are resolved it is expected that dial signaling will be expanded. Early applications are expected to be on routes having combined cable and radio facilities. On such routes it is felt that the radio facilities normally used should be equipped for dialing as well as the cable circuits.

Summary and Conclusions

It has been shown that there has been a remarkable growth in transoceanic telephone communications in recent years. This has been stimulated particularly by the provision of underseas telephone cable circuits which provide service in quantity and quality far surpassing what has heretofore been available.

In contrast to the trend toward using

cable in long transoceanic circuits, for which radio has in the past been the means of telephone communication, there is an opposite trend toward using radio for domestic networks and the rather short links to near-by islands. The expanding use of line-of-sight and over-the-horizon radio systems is clearly evident. These systems provide circuits in quality and quantity comparable to what can be provided by wire methods.

High-frequency ionospheric reflection radio continues to be a major method of transoceanic communications. The extension of single-sideband techniques to those cases where this method was not already used and the provision of higher power is improving these communications.

As the major telephone networks are brought closer together by these new facilities, a need exists for improved operating methods including dial signaling. Means have been developed to apply these methods to high-frequency radiotelephone circuits in spite of the transmission difficulties due to multipath fading.

In conclusion, the continuation of present trends seems a certainty. Expansion in the domestic telephone network by means of microwave radio will continue. Additional over-the-horizon radio systems and underseas telephone cable systems for public telephone service will be provided as the circumstances warrant.

The provision of high-grade telephone circuits by underseas cables requires large investments. To help offset the costs of such facilities, means are being developed to use existing cables more efficiently. For example, a system called TASI (Time Assignment Speech Interpolation) is under development in the Bell Telephone Laboratories and seems likely to double the capacity of existing cables. It will do this by taking advantage of the times when people are listening or pausing. The voice channels they leave temporarily unused will be automatically assigned to other talkers; and when a listener starts to talk, he will instantly have the use of a channel which another person has left idle. Wider bandwidth-repeatered underseas cables will also be developed to reduce the cost per circuit where large quantities of circuits are required. The recently announced cable system between the United Kingdom and Canada promises to be a step in this direction, providing 60 circuits with both directions of transmission in the same cable sheath. These developments and others will make for an ever-expanding and more economical Global Public Telephone Communication Network.

Design Parameters for Optimizing the Efficiency of Thermoelectric Generators Utilizing P-Type and N-Type Lead Telluride

R. W. FRITTS
NONMEMBER AIEE

THERMOELECTRIC generators are presently being developed for many uses which range from small flame-operated power supplies for automatic control systems to several kilowatt-size thermopiles for prototype evaluation of marine propulsion systems.[1] This renewal of interest in thermoelectric power generation and thermoelectric heat pumping is based on recent advances in the development of special semiconducting materials which exhibit properties intermediate to metals and semiconductors of the type used in *P-N* junction transistor devices. Often called "semimetals," these thermoelectric materials are prepared from semiconducting crystals which are more heavily doped than transistor devices to provide higher electrical conductivity. Such materials exhibit much lower thermal conductivity than do metals and substantially higher thermal emf (electromotive force).

Thermoelectric generators offer a number of advantages over d-c power sources, such as batteries and rotating machines. They are of simple structure requiring no moving parts; they are silent in operation and capable of virtually endless shelf life when used as emergency stand-by power sources. To what extent thermoelectric generators will replace conventional means for generating electricity will depend largely upon the efficiencies that can be obtained in the semimetal materials now under intensive study in many laboratories.

The basic unit of a thermoelectric generator is a thermocouple comprising a *P*-type element and an *N*-type element placed electrically in series and thermally in parallel as shown in Fig. 1. The *P*-type element conducts electricity through the transport of positive holes end when placed in a temperature gradient exhibits a positive voltage at the cold and with respect to the hot end. The *N*-type element conducts electricity through electron transport and exhibits a negative voltage at the cold end with respect to the hot end. The two voltages add to produce the net emf of the circuit. The metal electrodes serves simply as contact and mounting means and their electrical resistance can be ignored for the purposes of this discussion.

When the Seebeck coefficients of the respective materials, S_p and S_n, measured with respect to a reference metal such as copper, are known, the voltage developed

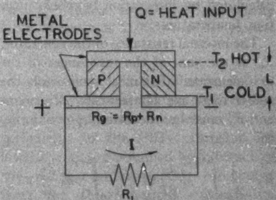


Fig. 1. Basic thermocouple circuit

Paper 59-909, recommended by the AIEE Solid State Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21-26, 1959. Manuscript submitted March 24, 1959; made available for printing April 28, 1959.

R. W. FRITTS is with the Minnesota Mining & Manufacturing Company, St. Paul, Minn.

The author wishes to acknowledge the contributions of R. E. Fredrick, William Huck, and Dr. J. D. Richards to the materials research program and the measurement of transport properties, and the discussions and guidance of Dr. Sebastian Karrer, who is singularly responsible for originating this research and development.

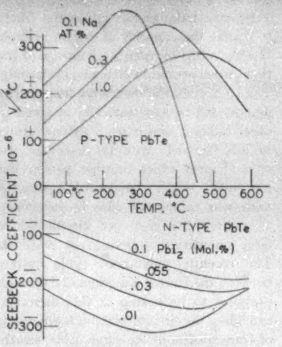


Fig. 2. PbTe Seebeck coefficient versus temperature data

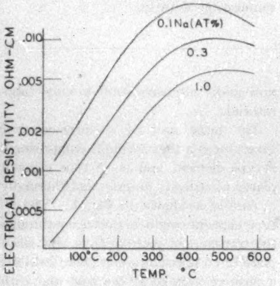


Fig. 3. Resistivity versus temperature for P-type PbTe alloys

in each element is obtained by integrating the Seebeck coefficient with respect to temperature between T_2 and T_1. The open-circuit voltage, termed the Seebeck emf, of the thermocouple is the sum of these two integrals:

$$V=\int_{T_1}^{T_2} S_p dT-\int_{T_1}^{T_2} S_n\, dT$$

In general, the Seebeck coefficients of the semimetals are quite temperature dependent and hence the voltages developed in the thermocouple elements must be calculated by integration between junction temperatures.

The internal resistance of the generator, R_g, is the sum of the resistances of the two elements. For most semimetals the resistivities are temperature dependent and R_g must be calculated by evaluating the integral

$$R_g=\frac{1}{A_p}\int_{T_1}^{T_2}\rho_p\frac{dx}{dT}\,dT+\frac{1}{A_n}\int_{T_1}^{T_2}\rho_n\frac{dx}{dT}\,dT$$

This integral expression for R_g can be evaluated when the resistivity temperature dependence is known and only when the temperature distributions within the semimetals are specified.

The power output of a specific thermocouple can be calculated from the foregoing integral expressions. The heat transferred through the thermocouple elements due to thermal conductivity under open-circuit conditions represents an irreversible loss and must be taken into account in calculating the efficiency of the thermocouple. The thermal conductivity of most semimetals is also temperature dependent and is the major factor controlling the temperature distribution within each element.

Ioffe[2] and others have treated the problem of thermoelectric efficiency thoroughly to find that the maximum efficiency is given by the expression

$$\eta_{\max}=\frac{\Delta T}{T}\,\frac{\sqrt{1+Z\dfrac{(T_2+T_1)}{2}}-1}{\sqrt{1+Z\dfrac{(T_2+T_1)}{2}}+T_1/T_2};$$

$$Z=\frac{(S_p-S_n)^2}{(\sqrt{K_n\rho_n}+\sqrt{K_p\rho_p})^2}$$

where Z is defined as the Figure of Merit and represents the material dependent parameters. S, K, and ρ refer to the Seebeck coefficient, thermal conductivity, and electrical resistivity of the respective elements. Under conditions of maximum efficiency the cross-section ratio for thermocouple elements of the same length is

$$\frac{A_n}{A_p}=\sqrt{\frac{K_p\rho_n}{K_n\rho_p}}$$

The Figure of Merit combines three temperature dependent parameters and hence it also is quite temperature dependent. This temperature dependence of the Figure of Merit must be taken into account if the preceding expression for operating efficiency is to be evaluated with accuracy. Where Z can be plotted against temperature from experimental data, the effective value of Z to be used in the efficiency calculation is the average value of Z obtained from

$$Z_{\text{ave.}}=\frac{1}{\Delta T}\int_{T_1}^{T_2} Z\, dT$$

Where P-type and N-type materials of identical properties are used the Figure of Merit can be reduced to

$$Z=\frac{S^2}{K\rho}$$

where S, K, ρ refer to the properties of either material alone. In the subsequent analysis the Figure of Merit of the

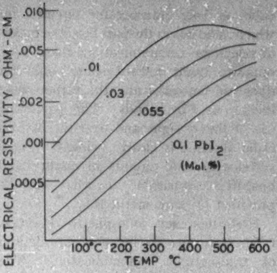


Fig. 4. Resistivity versus temperature for N-type PbTe alloys

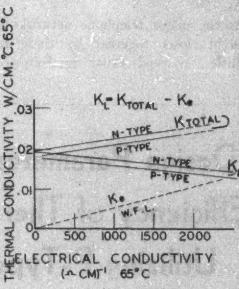


Fig. 5. PbTe thermal conductivity data determined for 65 C

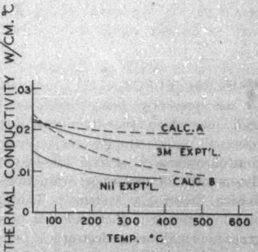


Fig. 6. Experimental and calculated thermal conductivity temperature dependence

N-type element and P-type element will be treated separately for the sake of simplicity and the average value of the two will be used to solve for the efficiency of the combination in a thermocouple.

Examination of the expression for maximum efficiency reveals that a high value of Figure of Merit is required for high efficiency; high hot junction temperatures and large temperature differentials

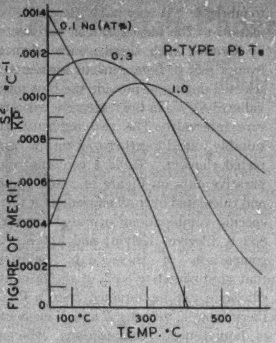


Fig. 7. Figure of Merit versus temperature for P-type PbTe alloys

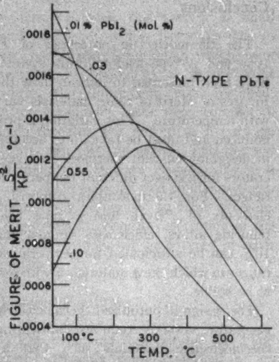


Fig. 8. Figure of Merit versus temperature for N-type PbTe alloys

also increase the efficiency. Hence, for optimum efficiency thermoelectric materials exhibiting high Seebeck coefficient, low electrical resistivity, and low thermal conductivity are desired. Efficiency can be improved by utilizing very stable chemical species for the thermocouple elements so that they may be operated at high hot junction temperatures if high values of $S^2/K\rho$ can be maintained at these high temperatures.

Because of the variation of Z with temperature for most semimetals, there will seldom be a specific homogeneous composition which will exhibit the optimum efficiency over all temperature ranges to be encountered in device design. It is the purpose of the next section to describe the properties of a typical thermoelectric material and present methods for selecting the best compositions for specific thermoelectric generator designs.

Thermoelectric Properties of Lead Telluride Systems

Lead telluride is the first of the new thermoelectric materials to be used extensively in operating thermoelectric generator devices. The successful operation of lead telluride generators at hot junction temperatures of 600 C (degrees centigrade) for several years demonstrates the chemical stability of the binary compound and proves the existence of compatible contacting and mounting schemes.

Lead telluride (PbTe) is a binary compound which melts at 922 C and its chemical binding is largely covalent. The material can be prepared in either *P*-type or *N*-type form and the carrier concentration of either type can be varied over about two orders of magnitude by suitable doping. The system exhibits high carrier mobility for electrons and positive holes, (about 1,800 and 880 cm^2/volt sec respectively at room temperature). Unlike silicon and germanium the mobilities are independent of carrier density up to concentrations of about 10^{19}/cc and the systems are effectively nondegenerate in this same range.

The thermoelectric properties of *P*-type and *N*-type PbTe alloys are shown in Figs. 2–6, plotted for the temperature range over which reproducibility can be achieved for long term service. At the lower temperatures the Seebeck coefficient and resistivity increase with increasing temperature characteristic of an extrinsic semiconductor. At elevated temperatures this trend reverses and further temperature increase induces intrinsic ionization of electrons and holes typical of a pure semiconductor. Impurity doping which increases the density of electrons or holes at all temperatures has a twofold effect on these alloys. Increased doping reduces both the Seebeck coefficient and the resistivity in the extrinsic range. Increased doping also extends to higher temperatures, the point at which the transition from extrinsic to intrinsic behavior occurs. This is most important since the highest values of Figure of Merit are obtained in systems which are primarily extrinsic in nature.

Fig. 5 shows how thermal conductivity is influenced by increased doping which causes a proportional increase in electrical conductivity. The upper curves specify the total thermal conductivity and the lower curves indicate how the lattice conduction is reduced by the addition of the doping agent to the PbTe lattice. This reduction in the lattice conduction is to be expected as the solid solution of the doping agent introduces scattering centers in the periodicity of the PbTe crystal. The lower curves are obtained by subtracting the nondegenerate electric component from the total thermal conductivity. This electronic component is given by the following:

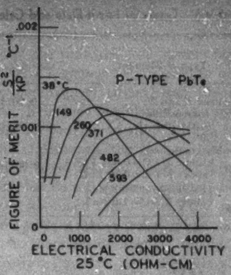


Fig. 9. Figure of Merit isotherms of P-type PbTe alloys plotted against room temperature conductivity

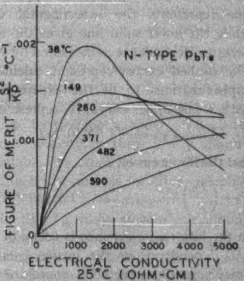


Fig. 10. Figure of Merit isotherms of N-type PbTe alloys plotted against room temperature conductivity

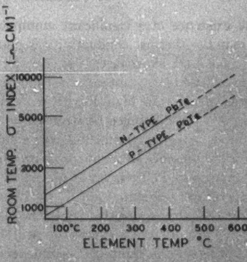


Fig. 11. Room temperature conductivity index to optimum $S^2/K\rho$ versus temperature

Table I. Figure of Merit Data and Calculated Efficiencies of Different PbTe Combinations

Type of Elements	N	P	Average	Calculated Efficiency
Idealized P and N	0.00133 C^{-1}	0.00105	0.00119	11.70%
Single composition N....0.055 PbI_2 P...1.0 Na	0.00115	0.000835	0.000992	10.28%
Doubly segmented N... below 371 C=0.055 PbI_2 above 371 C=0.10 PbI_2	0.00119		0.00109	
P... below 315 C=0.3 Na above 315 C=1.0 Na		0.00100		11.02%

Note: only $T_2=866$ K $T_1=311$ K $S^2/K\rho$ C^{-1}.

$$Ke=2\left(\frac{k}{e}\right)^2\sigma T=1.5\times10^{-8}\sigma T$$

The temperature dependence of thermal conductivity for a single PbTe composition is shown in Fig. 6. The upper solid line represents the experimental data while the lower solid line gives the temperature dependence as measured by Nii.[3] The dashed curves represent calculated approximations to fit the experimental data. The upper curve, A, represents

$$K_T=K_{(\text{lattice } T=65\text{C})}+1.5\times10^{-8}\sigma T$$

and the lower curve, B,

$$K_T=\left(\frac{To}{T}\right)K_{(\text{lattice } T=65\text{C})}+1.5\times10^{-8}\sigma T$$

In the latter case account is taken of the $1/T$ dependence of the lattice conductivity normally expected in crystals of this type. From theory the most realistic expression for thermal conductivity is

$$K_T=\left(\frac{To}{T}\right)K_{(\text{lattice } T=65\text{C})}+1.5\times10^{-8}\sigma T+\text{radiation term}$$

The existence of a significant ambipolar term in the thermal conductivity is questioned as high extrinsic carrier concentrations will tend to suppress the minority carrier density in high Figure of Merit alloys. The difference between curve B and the upper experimental curve suggests the presence of a radiation transfer mechanism in PbTe crystals placed in a temperature gradient.

For the purposes of this paper the first thermal conductivity expression will be used wherein the temperature independent lattice term is taken from the lower curves of Fig. 5. The plots of Fig. 5 are considered to be the most accurate data available on the thermal conductivity of doped PbTe.[4] The resulting error in the calculated Figures of Merit will be conservative and the true materials efficiencies will be somewhat higher than the values to be given here.

The values of $S^2/K\rho$ calculated from the data of Figs. 2 to 6 are plotted in Figs. 7 and 8 as point functions of temperature. For both P-type and N-type systems low concentrations of doping agents, and hence low carrier densities, provide high $S^2/K\rho$ values at low temperatures. High concentrations of doping agents provide high $S^2/K\rho$ values in the higher temperature ranges.

The Figure of Merit isotherms in Figs. 9 and 10 are obtained by cross-plotting the data of Figs. 7 and 8. If in turn the room temperature electrical conductivities corresponding to the $S^2/K\rho$ maxima of Figs. 9 and 10 are plotted against the isotherm temperatures, an index is obtained which specifies how the composition of the thermocouple element must be varied to provide the maximum value of $S^2/K\rho$ at each point in the temperature gradient. Room temperature electrical conductivity is a more convenient measure of the actual electron or hole concentration than would be the concentration of the doping impurity. The reason for this is that any one of several P-type or N-type doping agents may be used to alter the electrical properties in an approximately equivalent manner. Once the desired room temperature conductivity index is specified then the specific concentration of the selected doping agent may be obtained directly from the doping curves.[5]

When thermocouple elements are constructed in this manner against known temperature distributions, the highest mean value of $S^2/K\rho$ is obtained yielding the ultimate materials efficiency possible in combinations of lead telluride. This room temperature conductivity index is given in Fig. 11 for both P-type and N-type lead telluride alloys. The corresponding Figure of Merit variations with temperature of these optimized P-type and N-type thermocouple elements are represented by the envelopes to the curves of Figs. 7 and 8.

Thermocouple elements with the prescribed concentration gradient of doping agent for maximum $S^2/K\rho$ are difficult to fabricate. A reasonably close approximation to the ideal case can be effected by forming each thermocouple element from a pair of homogeneous segments of selected doping to provide the best mean value of $S^2/K\rho$ in the respective temperature intervals. The two segments are suitably joined together to form the completed element. Table I contains comparative data on Figure of Merit values and calculated over-all efficiency for three specific combinations operated between 866 K (degrees Kelvin) and 311 K. A comparison of these data indicates that approximately 85% of the ultimate efficiency can be obtained from a PbTe thermocouple utilizing P and N elements of homogeneous compositions, whereas the use of doubly segmented P and N elements affords about 95% of the ultimate value.

Conclusions

The thermoelectric properties of P-type and N-type lead telluride systems have been presented to indicate how the Figures of Merit of these materials vary with temperature. This variation is pronounced and must be taken into account in designing efficient thermoelectric generators for use over different temperature ranges. For thermocouples comprising elements of P-type and N-type lead telluride alloys, efficiencies approaching 10% can be anticipated for temperature ranges in which these materials are known to be stable.

The design of optimized thermoelectric generator systems requires fabricating thermoelectric materials such that at each point in a thermocouple element the crystal is doped to provide the maximum Figure of Merit for the temperature at that point. While the data presented herein pertain to lead telluride systems, the same analysis can be applied to other thermoelectric materials as well.

References

1. HIGH VOLTAGE SEMIMETAL THERMOCOUPLES, **R. W. Fritts.** *Electronic Design*, New York, N. Y., May 28, 1958, pp. 28–31.

2. THERMOELECTRIC GENERATORS, **A. F. Ioffe.** *Izvestiya Akademii Nauk SSSR*, Moscow, U.S.S.R., no. 1, Jan. 1956, pp. 76–80.

3. MEASUREMENT OF THE THERMAL CONDUCTIVITY IN SEMICONDUCTORS, **Riro Nii.** *Journal*, Physiological Society of Japan, Tokyo, Japan, vol. 13, 1958, pp. 769–70.

4. THERMAL CONDUCTIVITY OF PbTe DOPED WITH P-TYPE, N-TYPE AND ELECTRICALLY INERT SOLUTES, **W. V. Huck.** *Bulletin*, American Physical Society, New York, N. Y., paper FA4, series II, vol. 2, no. 4, Apr. 25, 1957.

5. THERMOELECTRIC GENERATORS, **R. W. Fritts, S. Karrer.** *U. S. Patent no. 2,811,571*, Oct. 29, 1957.

The Flagstaff-Phoenix TJ Radio System

JOHN W. HIDY
MEMBER AIEE

THE Bell Telephone Laboratories, Inc., have recently completed the development of a microwave radio system to operate in the 11-kmc (kilomegacycle) band. The system is coded *TJ* and is designed for short-haul and medium-haul systems. One of the first systems was installed between Flagstaff and Phoenix, Ariz. It is hoped that a description of the system may be of interest to other engineers in the communications industry.

This paper will describe the layout and some of the special engineering features of the Flagstaff-Phoenix *TJ* radio relay system. Because this system is small and packaged, it has been possible to use prefabricated buildings which were temporarily located at a central point where the equipment was installed. After equipment tests were completed, each repeater was moved as a package to its permanent location. This method of operation appears economically feasible and it is hoped that similar arrangements may be applied to other projects. The detailed technical description of the *TJ* radio system will be covered in other papers.

Present Facilities

The telephone facilities on the Flagstaff-to-Phoenix route consisted of an open-wire line equipped for various types of carrier. By 1956 these facilities had been developed to nearly their maximum capacity. Studies indicated the requirement for circuits to northern Arizona would necessitate either expansion of the present route or development of a new route. The open-wire line was old and had been constructed with light poles. It would not therefore carry the load of additional crossarms and wire. Since it would not have been economical to rebuild the old line, there remained two alternatives: to construct a new open-wire line arranged for O type carrier, or to develop a radio relay route.

Cost studies indicated that the new open-wire line would cost less initially, but that, for the circuit cross sections greater than about 60 2-way circuits, the radio with L carrier multiplex equipment would be more economical. Since the estimated circuit requirements through the 20-year engineering period were in excess of 400 channels, it was decided to develop a radio relay route.

Radio Systems Considered

The Western Electric Company *TD-2* microwave system operating in the 4-kmc band was available. However, the route which had been selected did not lie along existing wire lines. Since the *TD-2* system assumes that the order wire and control circuits will be routed over wire lines, the system would have required an excessive expenditure for the construction of open-wire circuits. A second radio system to accompany the *TD-2* was considered for the order wire and control circuits, but this also proved too expensive. It was expected that future legs to Prescott and Winslow would be required within the engineering period. The addition of these legs using the *TD-2* system would have necessitated using two frequency-modulation terminals at Mingus Mountain and two at Mount Elden. This would have added appreciably to the cost of the system.

The development of a second Western Electric microwave radio system coded *TJ* and designed to operate in the 11-kmc band was nearing completion. The development schedule indicated that the first bays could be shipped during the fourth quarter in 1957. Since the additional circuits were required for service during the summer of 1958, the *TJ* system was considered a good candidate for the proposed route.

A comparison of the *TD-2* and *TJ* microwave systems indicated that the *TJ* system could be installed more economically and that its capacity would be sufficient for the engineering interval. Two of the principal costs factors favoring the *TJ* system were a built-in order wire and control circuit, and the fact that frequency-modulation terminals would not be required at future branching points. Although the system had not been tried in service, information indicated that its transmission characteristics should approach that of the *TD-2* when used within the design limitations. It was therefore decided to use the *TJ* microwave system for the Flagstaff-Phoenix route.

The TJ Radio System

The *TJ* microwave system is designed to operate in 11-kmc frequency range and is intended to operate over distances up to about 200 miles. The system provides for up to six 2-way radio channels. It is operated with one-for-one frequency diversity to protect against outages caused by multipath fading. This leaves a net capacity of three 2-way channels available to the system. The radio channels are suitable for either telephone or television.

In the *TJ* radio system one antenna is used for both transmitting and receiving in each direction. The frequency interval between the regular and associated diversity channel is 240 mc. Since multipath fading is frequency selective, it is believed that this difference will be sufficient to guard against fades of both channels in a given section at the

Fig. 1. TJ transmitter and receiver bay

Paper 59-943, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 23, 1959; made available for printing May 4, 1959.

JOHN W. HIDY is with the Mountain States Telephone and Telegraph Company, Phoenix, Ariz.

A project such as the Flagstaff-Phoenix TJ microwave radio system represents the combined efforts of many more engineers than can be named here. The author wishes to express his appreciation for the counsel and assistance of various members of the operating and engineering department of the American Telephone and Telegraph Company, the Western Electric Company equipment engineering group, the Bell Telephone Laboratories, Inc., and other members of the engineering department of the Mountain States Telephone and Telegraph Company.

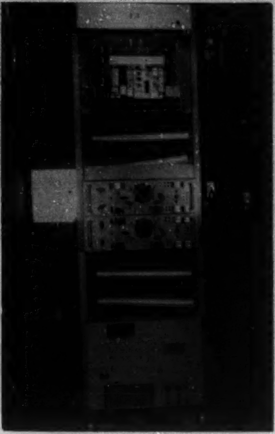


Fig. 2. D-1 alarm and control bay in TJ repeater station

same time. With all six channels in each direction equipped, there will be 80-mc separation between the unmodulated radio frequencies of each channel. Separation of each of the channels at the equipment is accomplished by means of waveguide band-pass filters. Additional segregation between the regular and diversity channels is obtained by means of vertical and horizontal polarization on the radio path. Fig. 1 shows a *TJ* transmitter and receiver bay.

With this arrangement, all transmitters and receivers in one direction are connected to the same antenna. The regular equipment bays and the diversity equipment bays for one direction are grouped and connected to separate waveguides, one for vertically polarized signals and the other for horizontally polarized signals. These waveguides are connected to a polarizer mounted on the rear of the antenna. The two polarizations are fed through a circular feed to the parabolic reflector. Thus, about a 30-db (decibel) discrimination between vertically and horizontally polarized signals may be obtained.

The base-band signals are fed to the regular and the diversity transmitters in parallel. At the receiving end the signals are received by two receivers and the strongest signal is selected by a diversity switch and fed to the next transmitter at a repeater point or to the multiplex equipment at a terminal. The diversity switch is designed to compare the level of the radio signals at the input of the receivers and the level of the 2,600-cycle control tones at the output of the receivers. If the level of the signals at the regular channel drops by a predetermined amount below the level of the signals of the diversity channel, a switch will result. The transfer switch has contacts arranged in a make-before-break sequence.

The design of the *TJ* system is such that the signals are returned to the base-band frequencies at the output of each receiver. This arrangement makes it practicable to apply the order wire, alarm, and control circuits directly to the base band without modulation. This is accomplished at each terminal or repeater by means of low-pass and high-pass filters to separate the voice band from the carrier band.

The order wire circuit is a 4-wire circuit so arranged that transmission is from the control terminal through each repeater to the far terminal and back to the control terminal. Each point is arranged to bridge the receiver on the outward direction of the circuit and the transmitter on the inward direction. Thus, for an attendant to talk to someone farther from the control terminal, he talks over a loop through the control terminal and out toward the remote terminal. The calling is over loud speakers. The order wire and control signals are transmitted over a 2,600-cycle tone using a standard single-frequency signaling unit. A 2,600-cycle slot filter is inserted in the voice path to make space for the signaling frequency. The 2,600-cycle tone is transmitted around the loop and back to the control terminal. Alarms are transmitted inward toward the control terminal by suitable coding applied to the tone. Also, orders are transmitted outward over the signaling tone. Up to 18 alarms may be received from each of seven radio stations. Also, ten orders may be transmitted to each station, five of which are used within the system. Fig. 2 shows a *D-1* alarm and control bay.

The order wire is so arranged that the failure of any station will permit a loop back at the next station toward the control terminal. This permits a check of all the stations between the control terminal and the station in the failure condition. It is therefore possible to locate the failure before a repairman is dispatched to clear the trouble.

Route

The route of the Flagstaff-Phoenix *TJ* radio system extends northward from Phoenix 133 miles. The route and profile are shown in Fig. 3.

The Phoenix terminal is located in the seventh floor penthouse of the Phoenix main building. The radio room is designed to provide the space required for all the terminal equipment which may be required in Phoenix within the next 10 years.

The first two repeater sites were located just off the Black Canyon Highway. The first, Desert Hills, is 26 miles north of Phoenix, just east of the highway. Black Mesa is located about five miles southeast of Sunset Point near where the road reaches the top of Black Canyon. The third site is on Mingus Mountain not far from the old mining town of Jerome. The fourth site is on Mount Elden located about 3.7 miles northeast of Flagstaff. The Flagstaff terminal

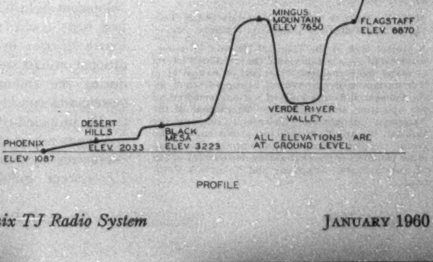


Fig. 3. Route and profile of Flagstaff-Phoenix radio relay route

Vol. Pt. 195

is located in the toll test room of the Flagstaff main building.

The path lengths are as follows:

Location	Miles
Phoenix Main to Desert Hills	25.88
Desert Hills to Black Mesa	21.09
Black Mesa to Mingus Mountain	38.55
Mingus Mountain to Mount Elden	47.84
Mount Elden to Flagstaff Main	3.72

The route extends from an elevation of a little over 1,000 feet to more than 9,000 feet. The extremes of temperature for the various sites range from over 115 F (degrees Fahrenheit) at Phoenix to more than 25 F below zero at Mount Elden.

There is nothing unusual about the first three paths of the route. The Mingus Mountain–Mount Elden path is in excess of 47 miles. However, the profile is such that many times the minimum clearance was obtained. The path from Mount Elden to Flagstaff doubles back on the Mingus Mountain–Mount Elden path with a little more than a 1-degree horizontal angle and about a 6-degree vertical angle. These two paths raised questions regarding rain attenuation and the frequency assignments of the two paths.

Tests made by the Bell Telephone Laboratories, Inc.,[1] indicated that an attenuation loss of about 1.3 to 1.4 db per mile may be expected when it is raining at the rate of one inch per hour. While heavy rains occur along the path, it is expected that they will generally be of a local nature and will not seriously affect the transmission. General rains will usually be of less density and should not cause difficulty even on the long path. At frequencies in the 11-kmc band, no attenuation due to snow or fog has been noted.

In the *TJ* system the frequencies are repeated every second section with protection from overreach interference obtained from cross polarization and zig-zagging. With the Mount Elden–Flagstaff path only slightly over 6 degrees from the Mingus Mountain–Mount Elden path, some concern was felt regarding the normal frequency assignment in the Mount Elden–Flagstaff section. To assure satisfactory operation on the initial installation, frequencies having the widest possible separation were selected for the two paths. It is hoped that experience will indicate that future radio channels may be assigned in their normal order which will make it possible to take additional radio channels into Flagstaff.

Multiplex Equipment

The initial radio channels of this system extend from Flagstaff to Phoenix and are designed for future legs to Prescott and Winslow. The circuit requirements are such that additions in groups of 12 circuits appear feasible. The length of the system together with these circuit requirements made it desirable to use an L-type carrier as the multiplex equipment. This is a single side-band suppressed carrier type of equipment which will load the system least of any of the channelizing equipments available. With the use of the L-type carrier, it is expected that up to 240 telephone circuits may be carried over a single radio channel as compared to 96 if *ON-2* carrier were used.

Towers

The antenna supports used on the Flagstaff-Phoenix route are self-supporting towers designed for the system. The towers at the repeater sites range in height from 60 to 120 feet. Extra height was provided at Mingus Mountain and Mount Elden to assure path clearance for the future Prescott and Winslow legs. The existing tower on the Phoenix main building was used. However, a special attachment was provided for mounting the antenna at the 25-foot point. No tower was required at the Flagstaff main building.

Antennas

The antennas selected for this system were 10-foot parabolic reflectors at all points except Flagstaff and Mount Elden toward Flagstaff. For this section, because it is very short, 5-foot antennas provided sufficient gain. A circular feed to the reflectors is used to accommodate both the vertically and horizontally polarized signals. The signals are fed into separate rectangular wave guides by a polarizer and a transducer mounted in the rear of the antenna. Two waveguides are therefore required for each direction of transmission. The gain of the 10-foot antenna is about 47 db, while that of the 5-foot antenna is 42 db.

The *TJ* radio system was designed for a medium-length light route service. With that objective some short cuts could be made in favor of economy. One of the major economies was obtained through the use of a-c powered equipment. This made the system subject to all of the variations which could occur at the end of long power transmission lines. Since this system was to form the backbone route to Northern Arizona, it would be used for dial, full-period leased wire, Teletype circuits, etc. Transmission of these services makes it imperative that interruptions from all sources be kept at a minimum. Since the power supply to any telephone equipment is one of the possible sources of service interruptions, it became necessary to select a power supply which was both reliable and relatively inexpensive.

Fig. 4. Power plant installed at a repeater station

Three types of emergency power plants were available which might have been used with the *TJ* radio system, as follows:

1. Storage battery plant without an engine stand-by.
2. A standard engine-driven alternator, automatically started.
3. A no-break type of gasoline engine stand-by power plant.

The storage battery stand-by plant would require large motor alternators for converting the d-c power to a-c power to satisfy the requirements of the *TJ* radio system. This not only required more floor space than was available, but with one exception it proved to to be too expensive for this route. The exception was at Phoenix Main, where there was enough capacity available in an existing *L-3* motor alternator set to power the *TJ* terminal.

Standard automatic-start engine alternators were available which might have been used for stand-by power. However, the design of these units permitted a power break of about one minute, which, with the time required for the klystrons to return to the correct frequency, resulted in a break of from 3 to 5 minutes. Since this was to be the main link to Northern Arizona, it was not desirable to expose the system to such interruptions in the power supply.

A no-break type of stand-by power

Fig. 5 (left). Building and power plant foundations ready for building

Fig. 6 (above). TJ repeater buildings on flatcars at Phoenix

plant was available which had been used with considerable success by other utilities near Phoenix. The machine consists of an a-c motor, an alternator, and a heavy flywheel on the same shaft. There is a suitably sized gasoline engine which may be coupled to the flywheel by a friction clutch. When the primary source of power fluctuates beyond predetermined limits, the a-c motor is disconnected from the power source and the clutch is engaged to start the engine. The constants of the machine are such that only a slight dip in the frequency, and no change in the voltage, occur at the instant the engine starts. This type of machine has the further advantage of isolating the equipment from a-c line surges which often occur at the end of a long power line.

The alternators of these sets are self-excited with the field current derived from a separate winding in the stator, which is connected to the field circuit through Magamp-regulated rectifiers. The terminal voltage is regulated by a magnetic amplifier which constantly compares the output voltage to two voltage reference diodes in series.

Tests using an oscilloscope indicate that the output of the alternator is very stable as to voltage and frequency when driven by either the motor or the engine. At the instant of transfer there is a small momentary drop of frequency and no observable change in the alternating voltage. A transient of about $2^1/_2$ volts peak-to-peak is introduced into the -400-volt output of the radio power supply. It may be noted that this figure does not exceed transients introduced by the normal operations of relays associated with the radio equipment. Fig. 4 shows the power plant installed at a repeater station.

Buildings

The *TJ* radio equipment was designed to be installed in small and compact buildings. When the system was being designed, thought was given to a building similar to the freight containers used jointly by the railroads and trucking companies. The building unit was required to be large enough to contain the power plant, the radio, and order wire equipment for a fully developed *TJ* radio route. It was designed to be handled with a moderately sized crane; its dimensions to meet the requirements for transportation on the highway.

Such buildings may be treated independently of the foundation. It was thus possible to pour six concrete piers to serve as a foundation. The piers were located at each corner of the building, and one on each side near the center. An additional pier was poured for an engine foundation. The piers were prepared by the contractor as a part of the work of preparing the tower foundations.

Fig. 5 shows the foundations for building and power plant.

With these thoughts in mind, arrangements were made with a manufacturer of highway freight trailers to construct four units, one for each of the repeater points. The units are $28^1/_2$ feet long and 8 feet wide. A partition is provided to separate the engine room from the equipment room. Each unit is equipped with two outside doors. Figs. 6 through 10 show these buildings in the various stages of the project.

Temporary Location of Buildings

When carrier or long radio systems are installed, a large amount of time is consumed in travel between lodgings and the repeater sites, and from site to site during the initial equipment tests and line-up. The *TJ* equipment and the buildings were of package design. This offered a good opportunity for overcoming some of the difficulties experienced with other systems. The buildings were therefore temporarily located in Phoenix where the equipment was installed and a large portion of the equipment tests made. While the buildings were still temporarily located, the first unit was connected to a temporary antenna, which was directed toward the Phoenix main building. The equipment in adjacent buildings was then connected together by waveguides in which 60-db attenuators had been inserted. This arrangement made it possible to transmit from Phoenix through each repeater station to the center of the Mount Elden equipment and back to Phoenix. It should be noted

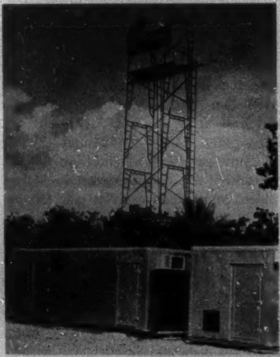


Fig. 7. Temporary antenna and two of the repeater buildings in contractor's yard at Phoenix

Vol.
Pt. 1
195

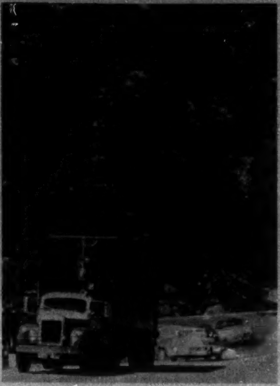


Fig. 8. View showing building fully equipped on truck en route to repeater site

that the circuit had to be looped at the alarm and control bay in the Mount Elden building and could not include the transmitters and receivers looking toward Flagstaff. It was possible to complete all the initial equipment tests at the repeater stations and to simulate transmission tests over all the route except the Mount Elden–Flagstaff section. The loop thus included a total of eight repeater sections.

Discussion of Installation at Central Point

It should be pointed out that the Flagstaff–Phoenix system was one of the early systems to be manufactured. While the prototype system had been given extensive trials and appeared satisfactory, the manufactured product had not been shaken down in the field. The installation of the equipment at a central point made it possible to complete the installation and to make many of the alterations with a minimum effort. In effect this amounted to doing four identical jobs within about 100 yards of each other.

In addition to the equipment installation work by the Western Electric Company, a considerable amount of work was required by an electrical contractor. Here again it was possible to do four identical jobs all in the same area. The result was a uniform installation of both equipment and electrical work.

To maintain the Flagstaff–Phoenix system, as planned, from three points (Flagstaff, Prescott, and Phoenix) it was necessary to train three maintenance groups. Even though this involved paying for maintenance for two groups during the training period in Phoenix, there was enough advantage in working in a central location to offset the additional cost.

Probably the biggest advantage to be gained from installing and testing is the great reduction in travel time to the various sites. It may be noted that work was being done in the buildings from May 19 to October 3, a period of 20 weeks or 98 working days. A conservative estimate of the travel time would be based on an average of two groups of two men each, driving an average of about 50 miles one way per day to reach the buildings, had they been placed initially on the sites. This equates to nearly 20,000 miles of driving and about 1,000 man-hours for travel time alone. In addition, the travel time would have been in excess of the regular working days or the working days would have been effectively shortened, which would have increased the installation interval.

While this system had an abnormally long installation and test-out period, it becomes evident that considerable savings may normally be effected by installing and testing the equipment in a central location.

Conclusions

The conclusions may be summarized as follows:

Installation Interval. By installing the equipment while the buildings were located on the construction lot, full advantage could be taken of the following items:

1. Travel time was reduced.
2. Each work operation was repeated in each repeater after the first.
3. It was unnecessary to pay maintenance for installation crews while working out of town.

While it is difficult to place dollar values on items such as the foregoing, the author estimates that the installation cost did not exceed one half of the cost of installing the equipment at the permanent sites.

Testing Efforts and Cost. The advantages of the centralized testing arrangement had all the advantages listed for installation plus the following:

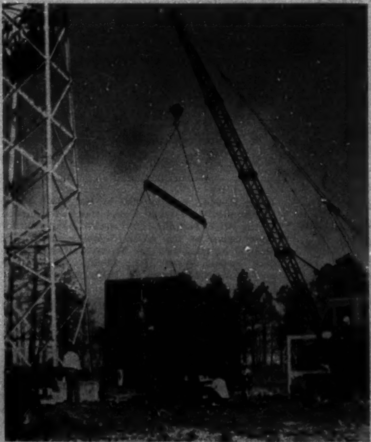


Fig. 9 (left). Building being placed on foundation by crane

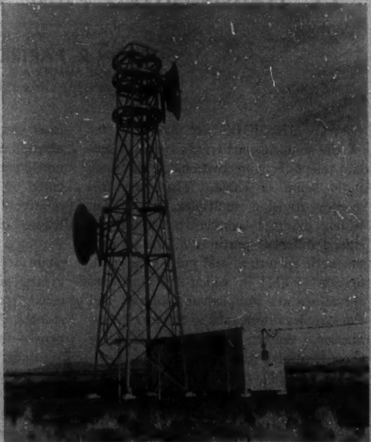


Fig. 10 (right). View of Black Mesa radio repeater station after completion

1. The benefits of a centralized training location.
2. A minimum transportation of testing equipment.
3. Travel time reduced to a minimum.
4. Possibility of obtaining a check of system performance with a minimum of effort.

It is estimated that the cost of the testing of the system did not exceed 75% of the cost of testing the same system, had the repeaters been placed in their permanent location initially. However, it should be recalled that this system was the first to be installed in buildings and then transported to the repeater site where it was retested before service. It is the author's belief that much of the latter testing may be eliminated on future projects, which would improve the 25% saving just noted.

No Trouble from Transportation. No equipment trouble was reported which could be attributed to the transportation to the sites.

No-Break Power Plants. The use of the no-break type of power plant appears to have considerable merit for installations such as the *TJ*. While it is not expected that the reliability will equal that of a storage battery plant, it should be adequate for this system. It is estimated that the interruptions due to loss of power will not exceed one per year for all five plants.

Possibility for Future Development. The method of installation described is, of course, most applicable to light route equipment which may be contained in small, easily transportable buildings. It is believed that the greatest advantage may be obtained by installing and testing the equipment at the factory. This arrangement would permit testing the radio bays and the *D-1* alarm and order wire bays as a system, which would eliminate wiring errors and defective equipment and insure a complete system being shipped. The factory wiring should include wiring for future radio channels.

After the installation and testing are completed, it is believed that the plug-in units should be identified with the associated bays and packed separately for shipment. This would give maximum protection to the equipment and would eliminate any difficulty from the hold-in screws, loosening and permitting the units to fall from the mountings during shipment.

An alternate arrangement of considerable merit would be to prewire the buildings for the electric and communication equipment at the factory where they are assembled. This arrangement has the disadvantage that the testing could not be completed at the factory. The actual installation effort on the job would be very small because of the plug-type connections used between the cable and the equipment. In this case it would probably be more economical to transport the buildings and equipment directly to the sites where the equipment installations would be completed.

Buildings. While the buildings on this project were satisfactory, it has been found that the temperature build-up resulting from the equipment heat is considerably more than was expected. The following suggestions may therefore be helpful on future projects:

1. The ventilation opening in the engine room should be covered with a filter.
2. The radio room of buildings in areas having ambient temperatures reaching 100 F and above should be air conditioned. The reason for this is that the heat dissipated for a fully equipped building would require an excessive volume of air to maintain suitable temperature for personnel and to guarantee save maximum working temperature for the equipment.
3. The radio room of buildings in areas having ambient temperatures ranging below 100 F should be equipped with forced ventilation controlled by a temperature switch. The air intake should be equipped with suitable filters and all openings covered with motor-operated shutters.
4. Changing the lighting from incandescent lighting (initially provided in the project) to fluorescent fixtures is suggested, to alleviate the condition of sharp and dense shadows on both the radio equipment and the test equipment.

Reference

1. RADIO ATTENUATION AT 11 KMC AND SOME IMPLICATIONS AFFECTING RELAY SYSTEM ENGINEERING, S. D. Hathaway, H. W. Evans. *Bell System Technical Journal*, New York, N. Y., vol. 38, 1959, pp. 73–97.

A Transistorized Pulse Code Repeater

GORDON R. PARTRIDGE
ASSOCIATE MEMBER AIEE

TWO METHODS are available for the simultaneous transmission of several telephone conversations on the same radio beam or cable. These are frequency division multiplex (FDM) in which different conversations are assigned different portions of the electromagnetic spectrum, and pulse time modulation (PTM) in which different conversations are sampled at intervals and the samples interleaved in time for transmission. Among the methods of pulse time multiplex are pulse position modulation (PPM), pulse width modulation (PWM), pulse amplitude modulation (PAM), and pulse code modulation (PCM). This last method is perhaps the most elegant of all, and the most complicated in terms of the necessary terminal equipment. The advantages procured by this increased complexity are a relatively high immunity to noise in the transmission system and, of special importance in some applications, the possibility of encrypting the signals for greater privacy by means other than the usual "scrambler." Under many circumstances, PCM systems will be less expensive and more reliable than other multiplex methods.

This paper describes an amplifier or "repeater" used for the transmission of PCM on cable. The repeater is one component of a complete communications system developed for the transmission of 24 simultaneous telephone conversations on military spiral-four cable. A brief summary of the system will be helpful in understanding the role of the repeater.

PCM System

A block diagram of the system is shown in Fig. 1. A number of telephone lines

---

Paper 59-944, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted March 19, 1959; made available for printing May 5, 1959.

GORDON R. PARTRIDGE is with the Raytheon Company, Waltham, Mass.

The author acknowledges with sincere appreciation the support and suggestions of a number of colleagues, in particular William L. Hatton of the Raytheon Company, and Martin Weinstock and Angelo Boniello of the U. S. Army Signal Research and Development Laboratories for whom the system was designed. The work described was performed under Signal Corps Contract *DA36-039-sc-73061*.

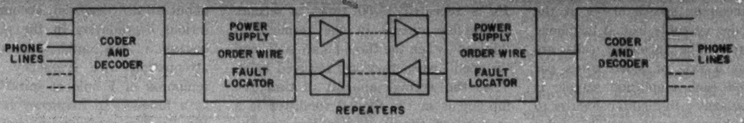

PHONE LINES
CODER AND DECODER
POWER SUPPLY
ORDER WIRE
FAULT LOCATOR
REPEATERS
POWER SUPPLY
ORDER WIRE
FAULT LOCATOR
CODER AND DECODER
PHONE LINES

Fig. 1. Block diagram of PCM system

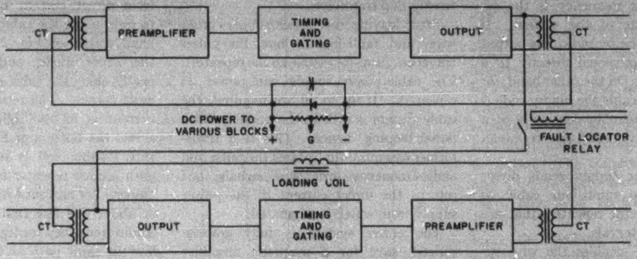

CT
PREAMPLIFIER
TIMING AND GATING
OUTPUT
CT
DC POWER TO VARIOUS BLOCKS
+
G
-
FAULT LOCATOR RELAY
LOADING COIL
CT
OUTPUT
TIMING AND GATING
PREAMPLIFIER
CT

Fig. 2. Block diagram of repeater

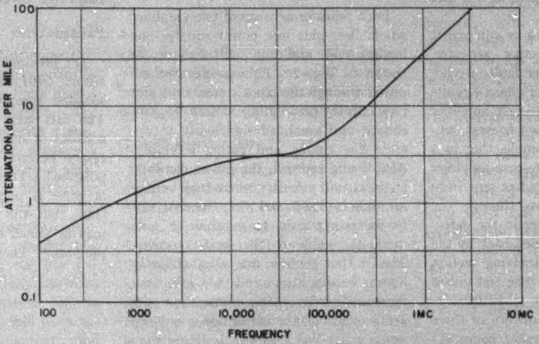


Fig. 3. Attenuation of spiral-four cable with frequency

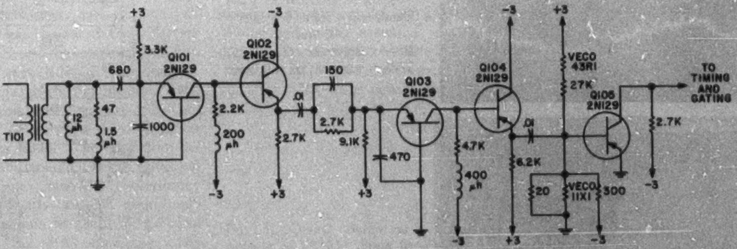

T101
12 μh
47
1.5 μh
680
+3
3.3K
1000
Q101 2N129
2.2K
200 μh
-3
Q102 2N129
-3
2.7K
+3
.01
150
2.7K
9.1K
+3
470
Q103 2N129
4.7K
400 μh
-3
Q104 2N129
-3
.01
6.2K
+3
+3
VECO 43R1
27K
20
VECO 11X1
300
-3
Q105 2N129
2.7K
-3
TO TIMING AND GATING

Fig. 4. Preamplifier

carrying incoming speech arrive at the coder-decoder assembly from a telephone switchboard. The incoming calls are sampled at a rate of 8,000 times per second, and each amplitude sample is coded by an analog-to-digital converter to a six-digit binary number. The pulse rate leaving the coder is therefore 24 times 8,000 samples per second times six digits per sample, or 1,152,000 bits per second. This frequency is the so-called "baud" rate of the system. If transmission is to be by radio, the output of the coder is connected directly to a radio transmitter. On the other hand, for transmission over cable the binary code is fed next to an assembly consisting of a cable-power supply, order-wire terminal, and fault locator.

The cable-power supply sends direct current down the spiral-four cable to provide power for the repeaters that are spaced at 1-mile intervals.

The order wire enables the wire supervisor to communicate by conventional audio-frequency signals with other attended points in the system by using the phantom circuit of the cable as a 2-wire line.

The fault locator contains equipment for identifying a defective repeater. In the event of a repeater fault, a relay within the dead repeater short-circuits the output of the west-to-east amplifier to the output of the east-to-west amplifier. Suppose, for example, that in a given repeater the west-to-east amplifier is defective. Then test pulses sent from the eastern terminal pass through the good east-to-west amplifier in the defective repeater, and are turned back by the relay to return to the receiving section of the eastern terminal. The test pulses consist of a special pattern, . . .101010*111*01010. . . in which the sequence of three *1*'s is repeated 1,000 times a second. A flip-flop circuit in the terminal is "flipped" when the three *1*'s are sent, and "flopped" when they are received back. Thus the average amount of time the flip-flop is in the flipped condition is a measure of the round-trip time to the faulted repeater. This time is indicated very simply by a D'Arsonval meter in series with the collector lead of one of the flip-flop transistors.

After leaving the power supply, order wire, and fault-locator box, the pulses travel on spiral-four cable to the repeaters. The cable-power supply can power 25 repeaters. If up to 50 are required, the cable system is split in the middle with a power-looping device. This is a transformer assembly that passes the pulse and audio-frequency order-wire signals, but returns the direct current to the power supply from which it originated.

The entire system is fully transistorized, save for a magnetic amplifier in one of the power supplies.

Introduction to the Repeater

Each repeater consists of two amplifier assemblies, plus one power supply, one loading coil, and one fault locator, as shown in Fig. 2. Pulses enter the repeater through the input transformer and pass into the preamplifier. Here they are shaped and amplified sufficiently to operate the timing and gating assembly. After being retimed, the pulses continue to the output circuitry where their widths are standardized and they receive sufficient power gain to continue 1 mile down the cable to the next repeater. The output pulses are approximately 3 volts in amplitude and 0.44 μsec (microsecond) wide. After propagation over 1 mile of cable, the amplitude is reduced about 40 db (decibels), accompanied by serious degradation in the pulse shape because of the rapid increase in cable attenuation with frequency. The attenuation of 1 mile of spiral-four cable as a function of frequency is shown in Fig. 3. The characteristic impedance at the baud frequency is approximately 120 ohms.

Power for the repeaters is obtained from direct current traveling from left to right along the cable in the upper left-hand corner of Fig. 2, passing through the Zener diode, and leaving the repeater along the cable at the upper right-hand corner. The return path for direct current is in the right-to-left direction across the bottom of Fig. 2. The order-wire loading coil is in series with this path. Each repeater takes 67 ma (milliamperes) of current with a drop of 6 volts, or about 400 mw (milliwatts) of power. Audio-frequency order-wire signals follow the same path as the direct current. The relay for looping back pulses during fault locating is also shown in Fig. 2.

Repeater Circuitry

PREAMPLIFIER

The preamplifier consists of five transistors, *Q101* through *Q105*, and their associated components, shown in Fig. 4. The first step in processing the pulses consists of differentiation. This is necessary because spiral-four cable is not polarized. The identity of a given pair of wires is preserved through a reel of cable, but the two wires within a *pair* may be joined to the cable connectors at random. Therefore, the preamplifier

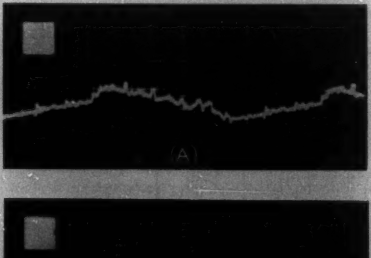
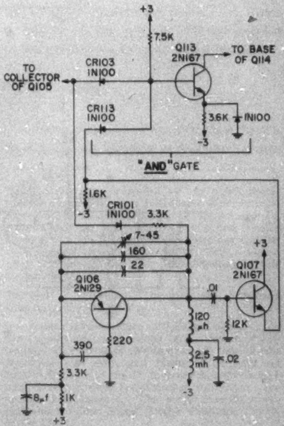


Fig. 5 (left). Waveforms. Scale shown by solid white square

A—Across secondary of input transformer; scale: 0.5 volt by 1.0 μsec

B—At collector of Q105; scale: 1.0 volt by 1.0 μsec

Fig. 6 (right). Timing and gating assembly



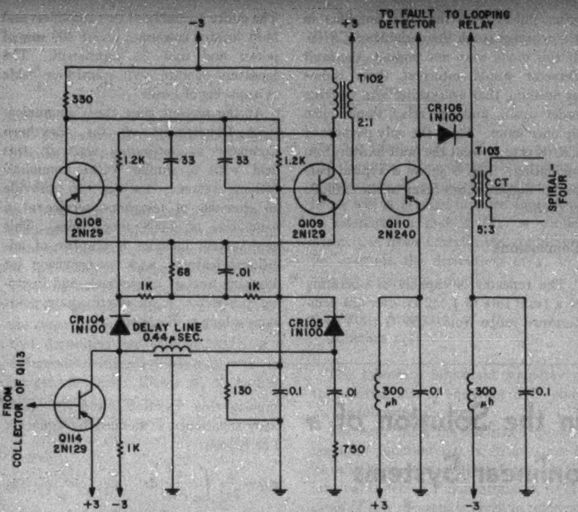


Fig. 7. Output circuitry

must be prepared to handle incoming pulses of either positive or negative polarity. After differentiation, each pulse has a positive and a negative portion. For example, an incoming pulse with a positive-going leading edge will produce a small positive pulse followed by a small negative pulse as a result of differentiating. Another benefit resulting from differentiation is the almost complete elimination of low-frequency components. This is necessary because the low frequencies are attenuated much less than the higher ones. Some patterns, such as ten *1*'s followed by ten *0*'s will have appreciable low-frequency content, and this may saturate the preamplifier if it is not removed. Differentiation is accomplished by the 12 μh (microhenrys) inductor shunted across the secondary of the input transformer.

The load impedances at the collectors of *Q101*, *Q103*, and the interstage coupling network between *Q102*, *Q103* are designed to provide increasing amplification with frequency, compensating more-or-less accurately for the shape of the attenuation curve of the cable. The capacitors between the emitters of *Q101* and *Q103* and ground help to remove high-frequency noise components without appreciably reducing the gain at the 1,152-kc baud frequency. Fig. 5(A) shows the waveform at the input of the repeater, and the amplified signal at the collector of *Q105* is shown in Fig. 5(B). This latter waveform includes a sinusoidal component that feeds through from the timing assembly when no pulses are present.

Timing and Gating

One of the largest problems in a PCM repeater is that of retiming the pulses. A certain amount of crosstalk occurs from pulse to pulse because exponentially-decaying tails are left by pulses passing through a high-pass network. The best remedy would be to direct couple the entire system, from the coder at one end to the decoder at the other. Since this is not practical, the next best thing is to retime the pulses so that the effects of pulse crosstalk (commonly called jitter) are not allowed to accumulate.

Retiming is accomplished by gating a timing pulse through an AND gate. The timing and gating system is shown in Fig. 6. The AND gate consists of diodes *CR103*, *CR113*, and the base-to-emitter diode of *Q113*. The gate is enabled by pulses at the collector of *Q105*, but the actual pulses passing through the gate and eventually appearing at the collector of *Q113* are from the oscillator *Q106*.

Transistor *Q106* and its associated components form a Hartley oscillator operating at the baud frequency of 1,152 kc. The oscillator is synchronized with the pulses by energy passing through diode *CR101* when the anode of this diode is more positive than −3 volts. The sine wave at the collector of *Q106* swings between the levels of −6 and 0 volts. When the waveform is more negative than −3 volts, energy feeds backwards from the oscillator through *CR101* to the collector of *Q105*. This accounts for the sine wave appearing in Fig. 5(B) during times when there are no pulses. The coupling between *Q105* and *Q106* is fairly critical. If it is too tight, any jitter at *Q105* is fed right through the timing system. If it is too loose, the oscillator may fail to keep in synchronism with the baud frequency.

Transistor *Q107* is an emitter follower for decoupling the oscillator from the gate.

Variation of the oscillator's natural frequency with temperature is reduced by the 160 μμf (micromicrofarad) capacitor, which is a temperature-compensating type selected to match the changes of the other circuit components. The over-all frequency change with temperature is about 5 kc, measured with the oscillator in a free-running condition and *CR101* disconnected. This is a small enough variation so that energy from *Q105* can hold the oscillator in synchronism over a temperature range from −55 C (degrees centigrade) to +65 C.

Output Circuits

Retimed pulses from the collector of *Q113* are fed into a grounded emitter amplifier for voltage gain, and thence to a flip-flop where the pulse width is standardized. This assembly is shown in Fig. 7. Transistor *Q109* is normally off and *Q108* is normally on. When a positive pulse from diode *CR104* reaches the base of *Q108*, transistor *Q109* is turned on and energy flows to the base of

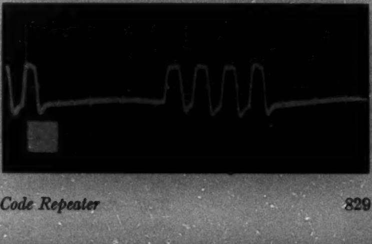


Fig. 8. Output pulses at cathode of CR106. Scale is shown by solid white square: 5.0 volts by 1.0 μsec

the power output stage *Q110* through transformer *T102*. After a delay of 0.44 μsec, the pulse that passes through the delay line turns *Q109* off again. In this manner, the width of the output pulses is standardized.

Transistor *Q110* is connected to its output transformer through a diode *CR106*. This diode is part of the fault-locating system. If pulses appear with reasonable regularity at the anode of *CR106*, a bias is maintained on a transistor in the fault-locating circuit (not shown) that keeps the fault-locator relay in the open-circuited state. Failure of pulses at the anode of *CR106* permits the relay to make contact. When this happens, pulses from the good amplifier in the repeater reach the cathode of *CR106*. If this diode were not present, the fault detector would interpret these pulses as meaning that everything is in working order again, and the relay would open up once more. Thus, the only purpose of *CR106* is to prevent the fault locator from oscillating. Fig. 8 shows a typical pattern of output pulses taken at the cathode of *CR106*.

Conclusions

The repeater is capable of operating at a baud rate of 1,152 kc over the temperature range from −55 C to +65 C. The entire assembly of two amplifiers and fault locator consumes about 400 mw of power and uses 26 transistors. The repeaters operate with spiral-four cable at a spacing of 1 mile.

At the present time the Communications Department of the Raytheon Company is continuing work in this field with a similar PCM communications system designed to provide 96 channels of telephone service at a baud rate of 4.608 megacycles. This new system includes a number of additional features, such as provision for dropping one set of channels and inserting a new set at an intermediate point along a long-haul system.

Recurrence Relations in the Solution of a Certain Class of Nonlinear Systems

ALFRED A. WOLF
MEMBER AIEE

THIS PAPER (a modified version of the material in an unpublished paper of the same name by the author presented at the AIEE 1957 Summer General Meeting) introduces aspects of a generalized mathematical theory for the analysis of a certain class of nonlinear systems. (Since the original presentation of this paper other work has appeared,[1-6] which enlarges on the comments made here and extends the present work.) It is shown that under a very broad class of conditions, consistent with physical realization, a class of systems exists for which it is possible to develop exact and unique mathematical solutions. These conditions are important since nonlinear differential equations do not in general possess such unique properties although linear differential equations do.

Those systems which are describable by ordinary nonlinear differential equations will be considered, such as:

$$Z(D)x(t)+\{x, \dot{x}, \ldots, x^{(m)}, \ldots\}=g(t) \qquad (1)$$

where $Z(D)$ is a linear integro-differential operator; $F\{x,\dot{x},\ldots,x^{(m)},\ldots\}$ is a nonlinear function of the response and its derivatives; and $g(t)$ is a forcing function. The nonlinear function shall be denoted simply as $F\{x\}$ in this paper and the remarks confined to nonlinearities which are functions of x only. Note that the method is not so limited.[1,2]

Existence and Uniqueness of Solution

In this section certain theorems concerning the existence and uniqueness of solution of a certain class of nonlinear systems are discussed. These theorems are most easily developed by the method of Picard iteration.[1,6,7]

Consider a sequence $x_0, x_1, x_2, \ldots, x_n, \ldots$, each of which are functions of time generated according to the following equations:

$$x_0(t)=\int_0^t g(\tau)y(t-\tau)d\tau \qquad (2)$$

$$x_n(t)=x_0(t)-\int_0^t y(t-\tau)\times F\{x_{n-1}(\tau)\}d\tau;\ n\neq 0 \qquad (3)$$

where $F\{x\}$ and $b(t)$ are known, single valued functions and $y(t)$ is a time function defined next. A limit

$$x(t)=\lim_{n\to\infty} x_n(t) \qquad (4)$$

exists such that $x(t)$ is the solution of a particular class of ordinary nonlinear differential equations provided $F\{x(t)\}$, $y(t)$, and $g(t)$ satisfy certain conditions. (The proof that $x(t)$ defined in equation 4 is a solution of equation 5 is given in Appendix I.) These differential equations may be written as follows:

$$Z(D)x(t)-F\{x(t)\}=g(t) \qquad (5)$$

where $Z(D)$ is a polynomial in the differential time operator D and $y(t)$ previously mentioned is defined by equation 6 as follows:

$$y(t)=\frac{1}{2\pi j}\int_{B_r}\frac{e^{st}}{Z(s)}ds \qquad (6)$$

where B_r is the appropriate Wagner-Bromwich contour in the complex s-plane enclosing the zeros of $Z(s)$.

Not only is equation 4 a solution of equation 5 but, under the conditions specified, it is the unique solution (the demonstration that $x(t)$ is the unique solution is given in Appendix II); and furthermore, it follows that all members of the sequence, though not necessarily the limit, are of exponential type (proof of this statement is demonstrated in Appendix III) converging uniformly to $x(t)$.

The linear operator $Z(D)$ is of the form

$$Z(D)=\sum_{n=N_1}^{N_2} f_nD^n \qquad (7)$$

Paper 59-919, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting and Air Transportation Conference, Seattle, Wash., June 21–26, 1959. Manuscript submitted October 11, 1957; made available for printing May 12, 1959.

ALFRED A. WOLF is with the Stromberg Carlson Company, a division of General Dynamics, Rochester, N. Y., and was formerly with the University of Pennsylvania, Philadelphia, Pa.

The author is greatly indebted to Dr. Harry Sohon, whose critical reading of the entire paper is much appreciated. He made many useful and pertinent suggestions concerning the body and format of the paper.

The author would like to express his deep appreciation to Dr. Y. H. Ku, who has been a source of constant encouragement and whose advice and guidance have been invaluable. Thanks are due to Dr. J. G. Brainerd and also to Mr. K. Lord of the Stromberg Carlson Company for their support.

The research reported in this paper and its final preparation has been made possible through support and sponsorship extended by the National Science Foundation under Research Grant NSF-G3673.

where

$$D=\frac{d}{dt}; \quad D^n=\frac{d^n}{dt^n} \tag{8}$$

and the coefficients f_n are restricted such that $y(t)$ defined in equation 6 is required to satisfy the condition that

$$y(t)=0(e^{\alpha_1 t}); \; t\geq 0 \tag{9}$$

in which α_1 is a real constant. The coefficients f_n are either constants or functions of time and s is a complex variable in the complex frequency plane corresponding to the differential time operator D. The bounds on the sum of equation 7 are constants which are positive or negative depending on whether $Z(D)$ is a differential, integro-differential, or an integro-operator. Either N_1 or N_2 may also be zero but not both.

The forcing function, $b(t)$, is required to satisfy a condition like equation 9 namely

$$g(t)=0(e^{\alpha_2 t}); \; t\geq 0 \tag{10}$$

where α_2 is a real constant.

Restrictions

A class of nonlinear systems is now defined by the following conditions imposed on $F\{x\}$, $y(t)$, and $g(t)$ which are consistent with physical realizability.

1. $F\{x\}$ is a single valued and uniformly continuous nonlinear function of x and its possible derivatives in some defined region. $F(x)$ is generally written to mean $F(x)=F(x, \dot{x}, \ldots)$.
2. Given a pair of functions $v(t)$ and $w(t)$ which are continuous in t or asymptotically like $e^{-\alpha_0 t}$, $\alpha_0>0$; then an $M>0$ and $a\geq 0$ exist such that

   $$|F\{v(t)\}-F\{w(t)\}|<Me^{-at}|v(t)-w(t)|$$

   is satisfied for all $t\geq 0$.
3. The function $y(t)$ is of exponential type.
4. The function $g(t)$ is of exponential type.

The latter two restrictions are already contained in equations 9 and 10 given previously. For convenience, $F(x)$ will be called a functional, provided restrictions 1 and 2 are satisfied.

The Power Function as a Functional

Consider the case when

$$F(x)=x^k \tag{11}$$

where $k\geq 2$. If

$$v(t)=A_1e^{-a_1 t} \tag{12}$$

and

$$w(t)=A_2e^{-a_2 t} \tag{13}$$

then equation 11 is a functional if $\alpha_1>0$ and $\alpha_2>0$. Since k in equation 11 is arbitrary then,

$$F\{x(t)\}=\sum_{k=2}^{M} a_k x^k(t) \tag{14}$$

also is a functional.

A function like equation 14 covers a broad class of nonlinear functions and hence will be studied in some detail here. In references 1 and 2, $F(x)$ is studied when it is more complicated containing, for example, the derivatives of x, in a manner similar to equation 14.

Analytic Analysis of the Solution x(t)

The following important property of the solution $x(t)$ of equation 5 is given in the following theorem.

Theorem: Under the conditions specified the limit function $x(t)$ is, in addition to its other properties, analytic for $t\geq 0$ in the interval in which F is continuous. (The proof of this theorem is given in reference 1 and follows from the theorem that the iterants $\{x_n\}$ under the conditions specified are all of exponential type as shown in Appendix III.)

Recursion Formulas

In view of the restrictions mentioned previously it is evident from the results of Appendix II that $x(t)$ is the unique solution of equation 5. Consequently, one is justified in developing the following recurrence relations. Indeed it can be shown[1] that in the absence of conditions like those in the section on restrictions of this paper, the results that follow would be false. Therefore, the recurrence equations which are developed as specified are consistent with the previously mentioned restrictions.

First consider the nonlinear function as defined by equation 14. Then expand the forcing function $g(t)$ as follows:

$$g(t)=\sum_{n=0}^{\infty} b_n t^n \tag{15}$$

where the b_n's are the Taylor coefficients of $g(t)$. This expansion is justified since $g(t)$ satisfies condition 4 and hence is analytic.

It is well known that if a function $x(t)$ is analytic, so is any integral power of the function analytic. Hence from the previous section it is seen that equation 14 is also analytic. Furthermore, if two functions are analytic then their difference is analytic.

Let $A(t)$ define an auxiliary forcing function which is single valued and differentiable associated with equation 5 according to the following relation

$$A(t)=g(t)-F(x) \tag{16}$$

From these remarks it is at once seen that $A(t)$ is analytic. Partitioning equation 5 to the right of the linear part and noting equation 16 the following equation results.

$$Z(D)x(t)=A(t) \tag{17}$$

If $A(t)$ were known then equation 17 could be solved explicitly for $x(t)$. The question of how to determine $A(t)$ in general is treated in detail elsewhere.[1] Note since $A(t)$ is analytic, it can be written as the power series which follows.

$$A(t)=\sum_{n=0}^{\infty} C_n t^n \tag{18}$$

Therefore, the solution of equation 5 under these conditions is

$$x(t)=\sum_{n=0}^{\infty} C_n Q_n(t) \tag{19}$$

in which $Q_n(t)$, given by equation 20 are the moments of the folded impulsive response of the linear part of equation 5

$$Q_n(t)=\int_0^t \tau^n y(t-\tau)d\tau \tag{20}$$

and

$$C_n=b_n-\sum_{k=2}^{M} C_n^{(k)} a_k \tag{21}$$

for zero initial conditions, and where $C_n^{(k)}$ is defined as follows:

$$\left[\sum_{n=0}^{\infty} C_n Q_n(t)\right]^k \equiv \sum_{n=0}^{\infty} C_n^{(k)} t^n \tag{22}$$

The $C_n^{(k)}$'s may be evaluated by the following method:

$$C_n^{(k)}=\frac{1}{n!}\frac{d^n}{dt^n}\left[\sum_{n=0}^{\infty} C_n Q_n(t)\right]^k\Bigg|_{t=0} \tag{23}$$

from which it is always possible to obtain $C_n^{(k)}$. It is noticed that when $Q_n(t)$ is a power function, equations 24 to 28 apply. [Since $Q_n(t)$ is an analytic function, this function can always be expanded into a power series if it is not already a power function. This will mean that the C's under the summation signs will themselves be power series. For example, suppose $Q_n(t)=e^{-nt}$, then $Q_n(t)=\sum_{s=0}^{\infty}(-1)^s n^s t^s/s!$ from which $C_n=\sum_{s=0}^{\infty}(-1)^s n^s/s!$. In such cases, it is simpler to use equation 23, or it is possible to develop recurrent relations to fit this case.]

$$C_n^{(0)}=\begin{cases}1; & \text{for } n=0\\ 0; & \text{for } n\neq 0\end{cases} \tag{24}$$

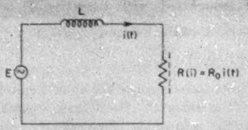


Fig. 1. **An inductive circuit L in series with a nonlinear resistor R(i)**

$$C_n^{(1)} = C_n \tag{25}$$

$$C_n^{(2)} = \sum_{r_1=0}^{n} C_{r_1} C_{n-r_1} \tag{26}$$

$$C_n^{(3)} = \sum_{r_2=0}^{n} \sum_{r_1=0}^{r_2} C_{r_1} C_{r_2-r_1} C_{n-r_2} \tag{27}$$

then

$$C_n^{(k)} = \sum_{r_k-1=0}^{n} \cdots \sum_{r_1=0}^{r_2} C_{r_1} C_{r_2-r_1} \cdots C_{n-r_{k-1}} \tag{28}$$

which can be used in simultaneous solution with equation 21.

It is worth while to point out that if the partition (generally this is a process of rearranging terms of equation 5 such that the left member is linear and contains the highest order derivative) is made at the highest order derivative then the moments $Q_n(t)$ will always be a power function.[1,2] It should also be remarked that the partition process must result in an equation such that the linear member contains the highest order derivative. Otherwise a recurrence relation will result which will require a knowledge of coefficients not yet found and moreover, the results will not converge in those cases where the solution is possible.

The Partition Process

The partition process consists essentially of the following steps satisfying the conditions previously mentioned.

1. The dynamic equation representing the system is partitioned so that the linear terms involving the highest order derivative are contained in one member of the equation.
2. A modified forcing function $A(t)$ is then obtained which consists of the actual forcing function, nonlinear, and possibly, some linear terms.
3. A proper choice is made of the expansion of $A(t)$ according to the particular properties of the linear, nonlinear, and driving functions. An auxiliary equation is thus formed.
4. The properly partitioned differential equation is then solved.
5. Certain coefficients which arise from the auxiliary equation are then calculated.
6. These coefficients, which appear in the partitioned equation, are then eliminated to obtain the exact solution.

An example is given to illustrate how to apply this technique to systems of one degree of freedom.[1]

Simple Application

As an illustration, consider a series circuit composed of an inductance with a nonlinear resistor $R(i)$ which is linearly dependent on the current through it and driven by a constant voltage E as shown in Fig. 1. The dynamic solution is

$$LD \cdot i(t) - R_0 i^2(t) = E \tag{29}$$

or

$$D \cdot i(t) - a i^2(t) = b \tag{30}$$

where

$$a = \frac{R_0}{L}$$

and

$$b = \frac{E}{L}$$

Assume that

$$i(0) = 0$$

Note on comparing equation 30 with equation 5 that $Z(D) = D$ and $F(i) = i^2(t)$ for this particular case. The solution is obtained either by comparing the constants of equation 30 with those of equation 5 and utilizing equations 19 and 21 or else the solution may be obtained directly utilizing the moment theorem given in equation 19 and the auxiliary equation given in equation 16. The latter approach is not only more direct but also is more general. One does not need to remember special formulas. It is only necessary to understand the preceding section. Therefore, the latter approach is given.

Since $Z(D) = D$ then $Z(s) = s$. On applying equation 6, $y(t)$ is recognized as the step function. That is,

$$u(t) = \begin{Bmatrix} 1; & t>0 \\ \frac{1}{2}; & t=0 \\ 0; & t<0 \end{Bmatrix} \text{ step function.}$$

$$y(t) = u(t) \tag{31}$$

and from equation 20

$$Q_n(t) = \frac{t^{n+1}}{n+1} \tag{32}$$

from which on substituting equation 32 into equation 19 gives

$$i(t) = \sum_{n=0}^{\infty} \frac{C_n}{n+1} t^{n+1} \tag{33}$$

The auxiliary equation is

$$\sum_{n=0}^{\infty} C_n t^n = b - a \left[\sum_{n=0}^{\infty} \frac{C_n}{n+1} t^{n+1} \right]^2 \tag{34}$$

Converting the multifold sum of the second member of equation 34 into a single sum yields

$$\sum_{n=0}^{\infty} C_n t^n = b - \sum_{n=0}^{\infty} C_n^{(2)} t^{n+2} \tag{35}$$

from which the nonlinear recurrence relation results:

$$C_n = bC_n^{(0)} - aC_{n-2}^{(2)} \tag{36}$$

Now noting that ($C_j = 0$ when $j<0$)

$$C_{n-2}^{(2)} = \sum_{k=0}^{n-2} \frac{C_k C_{n-k-2}}{(k-1)(n-k-1)} \tag{37}$$

yields

$$C_0^{(2)} = C_0^2$$

$$C_1^{(2)} = C_1 C_0$$

$$C_2^{(2)} = \frac{2}{3} C_0 C_2 + \frac{1}{4} C_1^2 \tag{38}$$

$$C_3^{(2)} = \frac{1}{2} C_0 C_3 + \frac{1}{3} C_1 C_2$$

When equation 38 is substituted into equation 36 the latter can be solved to give:

$$C_0 = b \tag{39A}$$

$$C_1 = 0 \tag{39B}$$

$$C_2 = -b \tag{39C}$$

$$C_3 = 0 \tag{39D}$$

$$C_4 = \frac{2}{3} b^3 a^2 \tag{39E}$$

$$C_5 = 0 \tag{39F}$$

. . .

Eliminating the coefficients C_n from equation 33 on substituting equation 39 yields the following expansion:

$$i(t) = bt - \frac{b^2 a}{3} t^3 + \frac{2}{3} \cdot \frac{b^3 a^2}{5} t^5 - \ldots \tag{40}$$

which is recognized as the expansion of

$$i(t) = \frac{b}{a} \tanh \sqrt{ab} t \tag{41}$$

This is the correct result as can be checked, since the differential equation 29 can be solved easily by separation of variables.

Discussion of Restrictions and Possibility of Generalization

In this paper certain restrictions were imposed upon terms appearing in a nonlinear differential equation. These restrictions define a class of nonlinear systems which have physical significance. In addition, these conditions simplify the demonstration that a particular unique solution exists. Although these restrictions are sufficient, it appears that they may not be necessary. Generally, an analysis that applies conditions that are both necessary and sufficient is considered a more desirable one. Hence, it might be well to inquire whether the restrictions can be relaxed without violating the conclusions drawn.

The two main restrictions imposed on the behavior of the nonlinear function F are quite independent of one another. On the one hand the nonlinear function was required to be a continuous function of the variable x and its various derivatives, if they were present, while on the other hand it was required to satisfy restriction 2. In the first place continuity of F is not necessary for the existence of a continuous solution. All that is really necessary for a continuous solution to exist is that the convolution integral of the functional and the folded impulsive response of the linear part exist and that the functional F be bounded by a function of the exponential type. It is easy to construct examples[1] which show the continuity of F is not sufficient to guarantee uniqueness of solution. Another condition is required. One such condition is the Lipschitz condition. This is a weak sufficiency condition. In this work, this condition has been suitably modified to enable proving certain theorems concerning stability of nonlinear systems.[1,5] This has resulted in restriction 2.

It is possible to substitute other conditions for 1 and 2 given here which are consistent with physical reality, without changing the uniqueness properties. Some work is now under way in this connection and may be the subject of a later paper.

Although the nonlinear term, $F(x)$ in the illustration, is a polynomial, corresponding results can be obtained when $F(x)$ is any function of x satisfying all the limitations of the section entitled, Restrictions. For example $F(x)$ may be a polynomial in the derivatives of x or combinations of polynomials of derivatives of x and polynomials in x. It should be remarked that if for any given t, the modified Lipschitz condition does not exist then the results of the fourth and fifth sections of the paper will generally fail. If for example

$$F\{x(t)\} = e^{-at}|x(t)|,\ a>0$$

then no unique power series expansion will exist at the origin if at $t=0$, $x=0$. However, it is possible to avoid such points by confining the analysis to points which do satisfy the restrictions.

It can be shown that other forms of the auxiliary function, $A(t)$ can be used. One such example would be the Dirichlet series.[1]

Solutions of nonlinear systems which form a complete orthogonal set can be treated in an interesting manner similar to that described.[1] It might be pointed out in this connection that a systematic treatment of analysis of nonlinear systems can be developed when $g(t)$ is a random function. Work in this area of stochastic processes in nonlinear systems belonging to the class described here will be reported on in the near future.[8]

Appendix I. Evaluation of the Limit of $x_n(t)$

This Appendix will outline the proof that $x(t)$ defined in equation 4 is a solution of equation 5. To do this, first consider equation 3. Subtracting the nth iterant from the $(n+1)$st iterant, taking absolute values, and applying the triangle inequality yields the result:

$$|x_{n+1}(t)-x_n(t)| \leq \int_0^t |F\{x_{n-1}(\tau)\} - F\{x_n(\tau)\}|\,|y(t-\tau)|\,d\tau \qquad (42)$$

From restriction 3

$$|y(t)| < Ke^{bt} \text{ for all } t>0 \text{ where the constants } K>0 \text{ and } b \text{ are real} \qquad (43)$$

Now applying equation 43 and restriction 2 to equation 42 yields:

$$|x_{n+1}(t)-x_n(t)| \leq MKe^{-bt} \times \int_0^t e^{-\beta\tau}|x_{n-1}(\tau)-x_n(\tau)|\,d\tau \qquad (44)$$

where $\beta \equiv a+b>0$, and $a>0$.

Since the iterants x_0 and x_1 are of exponential type, it is possible on referring to equation 3 to select a real positive constant q such that for all $t>0$,

$$\max|x_1(t)-x_0(t)| \leq q \qquad (45)$$

is satisfied. Then by the process of successive iterations of equation 44 one shows that for $n=1$:

$$|x_2(t)-x_1(t)| \leq q\frac{MKe^{bt}}{\beta}(1-e^{-\beta t}) \qquad (46)$$

for $n=2$

$$|x_3(t)-x_2(t)| \leq q\frac{M^2K^2e^{2bt}}{2\beta^2}(1-2e^{-\beta t}+e^{-2\beta t}) \qquad (47)$$

and continuing in this manner

$$|x_{n+1}(t)-x_n(t)| \leq q\frac{M^nK^ne^{nbt}}{n!\beta^n}(1-e^{-\beta t})^n \qquad (48)$$

It is evident that for a given $\epsilon>0$ an $N>0$ exists such that for all $n>N$ and $\beta>0$

$$|x_{n+1}(t)-x_n(t)| < \epsilon \qquad (49)$$

It is now desirable to evaluate the limit given in equation 4. Let $x_n(t)$ be written as the following identity:

$$x_n(t) \equiv x_0(t) + \sum_{r=1}^{n-1}[x_{r+1}(t)-x_r(t)] \qquad (50)$$

taking the limit of both sides as n

$$\lim_{n\to\infty}|x_n(t)| - |x_0(t)| \leq \lim_{n\to\infty}\sum_{r=0}^{n-1}|x_r(t)-x_{r-1}(t)| \leq \lim_{n\to\infty}\sum_{r=0}^{n-1} q\frac{S^n}{n!}$$

where

$$S \equiv \frac{MKe^{bt}}{\beta}(1-e^{-\beta t}) \qquad (51)$$

The right member is recognized as the series expansion of q_e^{S}. Then the left member is a uniformly convergent series by the Weierstrass M-test.[9] This shows the existence of the limit in equation 4. To show that the limit in equation 4 satisfies equation 5 one need only show that it satisfies

$$x(t)-x_0(t)-\int_0^t F\{x(\tau)\}\,y(t-\tau)d\tau \qquad (52)$$

which is the integral inversion of equation 5. The problem then is: Given $\epsilon>0$ arbitrarily, to show that there exists an $N>0$, such that for all $n>N$ and for all $t>0$:

$$\left|x(t)-[x_0(t)-\int_0^t F\{x_n(\tau)\}y(t-\tau)d\tau]\right| < \epsilon \qquad (53)$$

This will prove that $x(t)$ satisfies equation 52 identically. Writing the identity

$$x(t)-[x_0(t)-\int_0^t F\{x_n(\tau)\}y(t-\tau)d\tau] \equiv x(t)-[x_0(t)-\int_0^t F\{x_n(\tau)\}y(t-\tau)d\tau] \qquad (54)$$

and adding this to equation 3 gives, on simplifying, taking absolute values and applying the triangle inequality:

$$\left|x(t)-[x_0(t)-\int_0^t F\{x_n(\tau)\}y(t-\tau)d\tau]\right| \leq |x(t)-x_n(t)| + \int_0^t|F\{x_n(\tau)\} - F\{x_{n-1}(\tau)\}|\,|y(t-\tau)|\,d\tau \qquad (55)$$

Consider the second term of the right member of equation 55. Applying the condition of restriction 2 repeatedly for $n=0, 1, 2, \ldots$, etc., in a manner similar to that applied in the case of equation 44 yields

$$\int_0^t|F\{x_n(\tau)\}-F\{x_{n-1}(\tau)\}|\,|y(t-\tau)|\,d\tau \leq q\left(\frac{MK}{\beta}\right)^n\frac{(1-e^{-\beta t})^n}{n!} < \frac{\epsilon}{2} \qquad (56)$$

for all $t \geq 0$.

From equation 49 and noting the result of equation 51, it is evident that

$$|x(t)-x_n(t)| < \frac{\epsilon}{2} \text{ for all } t>0 \qquad (57)$$

Thus, equation 53 follows immediately and the limit in equation 4 satisfies equation 5.

Appendix II. Uniqueness of Solution

To complete the discussion of the properties of $x(t)$ it is shown that equation 4 is unique, that is, no other solution is possible.

Suppose some other solution does satisfy equation 5. Let this be designated as $x_a(t)$. Therefore, it too must satisfy equation 5 or equation 52 identically: Thus,

$$x_a(t) \equiv x_0(t) - \int_0^t F\{x_a(\tau)\}\, y(t-\tau)d\tau \qquad (58)$$

Subtracting equation 52 from equation 58 yields:

$$[x_a(t)-x(t)] = \int_0^t [F\{x(\tau)\} - F\{x_a(\tau)\}]y(t-\tau)d\tau \qquad (59)$$

Applying restriction 2 after taking absolute values results in:

$$|x_a(t)-x(t)| \leq M\int_0^t e^{-a\tau} \times |x_a(\tau)-x(\tau)|\,|y(t-\tau)|\,d\tau \qquad (60)$$

If c represents the maximum difference between the two solutions, and upon repeated iteration similar to that used to obtain equation 61, it is evident that

$$|x_a(t)-x(t)| \leq c\left(\frac{MK}{a}\right)^n \frac{(1-e^{-at})^n}{n!} \qquad (61)$$

for $t \geq 0$.

As $n \to \infty$ the right member approaches zero. Since the left member never can be negative, it follows that

$$x_a(t) \equiv x(t) \qquad (62)$$

Thus, there is one and only one solution.

Appendix III. Demonstration that the Iterants are of Exponential Type

Consider equation 63,

$$x_0(t) = \int_0^t g(\tau)y(t-\tau)d\tau \qquad (63)$$

The transform of equation 63 is

$$X_0(s) = G(s)Y(s) \qquad (64)$$

Since $g(t)$ and $y(t)$ are assumed to be of exponential type, and since $x_0(t)$ can be written as

$$x_0(t) = \frac{1}{2\pi j}\int_{B_r} G(s)Y(s)e^{st}ds \qquad (65)$$

It follows that $x_0(t)$ is of exponential type.[10,11] The first iterant is given by equation 66:

$$x_1(t) = x_0(t) - \int_0^t F\{x_0(\tau)\}\, y(t-\tau)d\tau \qquad (66)$$

$F\{x_0(t)\}$ is of exponential type[10] since the product of two or more functions of exponential type is of exponential type. Thus, $x_1(t)$ is of exponential type since sums of functions of exponential type are of exponential type. It may be concluded that $x_2(t)$ is of exponential type by reasoning similar to that employed in $x_1(t)$. Continuing in this manner, $x_n(t)$ is of exponential type.

Since it has been shown that

$$\lim_{n\to\infty} x_n(t) = x(t)$$

uniformly, it might be supposed that $x(t)$ is therefore also of exponential type. The fallacy in this reasoning lies in the fact that the limiting process may give rise to functions of higher order and not of exponential type. Numerous examples of this can be constructed.

References

1. A Mathematical Theory for the Analysis of a Class of Nonlinear Systems, A. A. Wolf. *Ph.D. Dissertation*, University of Pennsylvania, Philadelphia, Pa., June 1958.

2. Generalized Recurrence Relations in the Analysis of Physical Nonlinear Systems, A. A. Wolf. *AIEE CP 58-437* (available on request).

3. Analysis of Transcendental Nonlinear Systems, A. A. Wolf. *AIEE CP 58-995* (available on request).

4. Taylor-Cauchy Transforms for Solving A Certain Class of Nonlinear Systems, Y. H. Ku, A. A. Wolf, J. H. Dietz. *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 2, July 1959, pp. 49–61.

5. A Stability Criterion for Nonlinear Systems, Y. H. Ku, A. A. Wolf. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 78, July 1959, pp. 144–48.

6. Ordinary Differential Equations (book), E. L. Ince. Longmans, Green and Company, Inc., New York, N. Y., 1926; Dover Publications, Inc., New York, N. Y., 1956.

7. Existence Theorems (book), F. J. Murray, K. S. Miller. University Press, New York, N. Y., 1954.

8. Transform—Ensemble Method for Analysis of Linear and Nonlinear Systems with Random Inputs, Y. H. Ku, A. A. Wolf. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 15, 1960 (to be published).

9. Theory of Functions (book), K. Knopp. Dover Publications, Inc., 1945, p. 73.

10. Theory of Fourier Integral (book), E. C. Titschmarsh. Oxford University Press, London, England, 1937, p. 48.

11. Theorie und Undwendung der Laplace Transformation (book), G. Doetsch. Dover Publications, Inc., 1943, pp. 62–64.

12. Sur une Courbe, qui Remplit Toute une Aire Plane, G. Peano. *Mathematische Annalen*, Leipzig, Germany, vol. 37, 1890, p. 182.

13. Zur Theorie der Divergenten Reihen, O. Perron. *Ibid.*, vol. 76, 1915, p. 471.

14. Topics in the Theory of Functions of Several Variables, W. F. Osgood. *Monatshefte fur Mathematischer Physik*, vol. 9, 1898, p. 331.

15. On Isoclinal Lives, J. H. M. Wedderburn. *Proceedings*, Royal Society of Edinburgh, London, England, vol. 24, 1902, p. 400.

16. The Significance of Transients and Steady-State Behavior in Nonlinear Systems, A. A. Wolf. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 47, Oct. 1959, pp. 1785–86.

17. Approximations Sucessives et les Equations Differentielles, C. R. Cotton. *Comptes Rendus*, Academie de Sciences, Paris, France, vol. 140, 1905, p. 494; vol. 141, p. 177; vol. 146, 1908, pp. 274, 510; also *Mathematische Annalen*, vol. 31, 1908, p. 107.

18. Sur le Theoreme de M. Picard dans la Theorie des Fonctions Monogenes, E. Lindeloef. *Ibid.*, vol. 118, 1894, p. 454; also *Journal de Mathématiques et de Physique Appliqués*, Basel, Switzerland, vol. 10, no. 4, 1894, p. 117.

19. Traite d'Analyse, E. Picard. Vol. 3, p. 88; vol. 2, second edition, p. 340.

◆

Discussion

Thomas J. Higgins (University of Wisconsin, Madison, Wis.): This is an extremely interesting and well-written paper, whose context comprises account of a unique transform procedure for the solution of nonlinear systems. The discusser looks forward to account of further investigations, remarked in the last section of the paper as now in progress.

A. A. Wolf: The author wishes to convey his thanks to Professor Thomas J. Higgins for his very kind and encouraging remarks in his' discussion of this paper. May I also take this opportunity to thank all those people who were so kind to discuss my paper at the time of its presentation.

Since many people have asked about our work I have collected below a bibliography of both our published and planned work. I hope to resubmit references 2, 3, and 4. These were previously presented as conference papers. In the last section of the paper it was indicated that the present method could be extended to the case of nonlinear systems driven by random processes. Reference 5 deals with this subject and introduces the concept of the transform of a random process. In reference 6 we deal with a formal transform procedure of the method given here. This transform gives, in a formal way, a method of solving nonlinear problems comparable with the Laplace transform used in the solution of linear systems. In reference 7, I expect to deal with the general transform problem arising from the "partition theory." I also expect to present a paper at the forthcoming AIEE Winter General Meeting dealing with synthesis of nonlinear systems and its comparison to the Wiener-Lee-Bose theory.[8]

Among works scheduled for publication is a general paper[9] on nonlinear systems dealing with both deterministic and stochastic processes in analysis and synthesis of such systems. This was written in collaboration with Dr. Y. H. Ku.

BIBLIOGRAPHY

1. A Mathematical Theory for the Analysis of a Class of Nonlinear Systems, A. A. Wolf. *Doctoral Dissertation*, University of Pennsylvania, Philadelphia, Pa., June 1958.

2. Generalized Recurrence Relations in the Solution of Physical Nonlinear Systems, A. A. Wolf. *AIEE CP 58-437* (available on request).

3. Analysis of Transcendental Nonlinear Systems, A. A. Wolf. *AIEE CP 58-995* (available on request).

4. Transient Response of a Saturating Control System, J. H. Dietz, A. A. Wolf. *AIEE CP 58-295* (not available).

5. Transform—Ensemble Method for Analysis of Linear and Nonlinear Systems with Random Inputs, Y. H. Ku, A. A. Wolf. *Proceedings*, National Electronics Conference, Chicago, Ill., Nov. 1959 (to be pubished).

6. Taylor-Cauchy Transforms for Solving a Certain Class of Nonlinear Systems, Y. H. Ku, A. A. Wolf, J. H. Dietz. *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 2, July 1959, pp. 49–61.

7. On the General Theory of Transforms for Analysis of Nonlinear Systems Based on the Partition Theory, A. A. Wolf. *AIEE Transactions* (to be published).

8. Recent Advances in the Analysis of Nonlinear Systems, A. A. Wolf. To be presented at AIEE Winter General Meeting, Feb. 1960.

9. Theory of Nonlinear Systems (in 8 Parts), Y. H. Ku, A. A. Wolf. *Journal*, Franklin Institute, Philadelphia, Pa. (to be published).

10. A Stability Criterion for Nonlinear Systems, Y. H. Ku, A. A. Wolf. *AIEE Transactions*, pt. II (*Applications and Industry*) July 1959, pp. 144–47.

A Magnetic-Amplifier–Silicon-Transistor Power Supply for Missile Application

B. MOKRYTZKI
NONMEMBER AIEE

R. A. STUART
NONMEMBER AIEE

THE RECENT development of the first truly high-power silicon transistor and the demand for high-efficiency, regulated, solid-state power supplies have stimulated interest in silicon-transistor–magnetic-amplifier combinations. Apart from the usual ability of silicon devices to survive a higher temperature than their germanium counterparts, the new transistors, due to the high-voltage ratings and low internal losses, can control considerably more power than contemporary silicon units. These devices have proved to be extremely rugged and are virtually comparable with magnetic amplifiers in that respect. Thus, to all the advantages which have already been attributed to transistor-inverter–high-frequency magnetic-amplifier systems, it is now possible to attach the valuable qualities of silicon power transistors.

This paper describes in particular a solid-state power supply which is required to produce seven regulated direct voltages and two rms regulated alternating voltages. All the direct voltages are derived from the same high-frequency transistor inverter by providing several transformer secondaries and using a magnetic amplifier in each channel to regulate the average voltage output. The isolation afforded by the magnetic amplifiers provides capabilities for multichannel operation with only a single voltage reference. For the rms regulated outputs, inverters and magnetic amplifiers are again used along with a novel type of rms sensing circuit that provides a reliable and accurate measure of the rms value of a pulse-width modulated square wave. The resulting system is a high-performance power supply which is both reliable and efficient.

Problem and Over-all Approach

This power supply consists of three separate sections: section 1, the $B+$ supply, has seven independently regulated direct-output voltages for vacuum-tube plate supplies; section 2, the filament supply, delivers regulated a-c filament power to a total of 34 electronic tubes; section 3, the modulator motor supply, delivers regulated a-c power to a small motor. It is also desirable that each output be individually adjustable. The system is fed by two direct-current nominally 28-volt power supplies; input A is used to energize the filament supply only and has a voltage overshoot when it is turned on which may reach 45 volts for approximately 100 ms (milliseconds). Regulation of the filament supply output voltage is not necessary during this period, but the system must operate safely. Input B is turned on sometime after input A and supplies power for both the modulator motor supply and the $B+$ supply. The operating and environmental conditions are listed in the following and the output specifications are given in Fig. 1.

1. Input voltage: 28 volts d-c±10%.
2. Operating ambient temperature range: −40 C (degrees centigrade) to +100 C.
3. Operation at 100 *g* shock of 5.5-ms duration.
4. Operation at 100 *g* sustained acceleration for 20 minutes.
5. Vibration: 0.060 inch, peak to peak, from 10 to 55 cycles and constant 10 *g* acceleration from 55 to 2,000 cycles while operating.

The completed supply must be contained in two aluminum castings having a total volume of approximately 130 cubic inches. In spite of this small size, the over-all efficiency of the unit should approach 80%.

From an analysis of the various output and system requirements, several facts about the final assembly can be determined.

1. The operating ambient temperature range indicates that relatively high temperature components must be used, notably silicon semiconductor devices and wire-wound resistors.
2. The small size of the completed package and the requirement for an over-all efficiency approaching 80% suggests that a minimum number of components be used.
3. Since all the outputs except the modulator motor supply output must be regulated to ±1% of their nominal voltage and they should be individually adjustable, each output will probably have to be individually regulated.
4. The combination of a silicon transistor inverter with a magnetic amplifier offers an efficient and rugged method for deriving each of the required outputs. In order to regulate an input variation of ±10%, the magnetic amplifier need only support a portion of the output voltage, thereby insuring small size and satisfactory efficiency.
5. The total amount of d-c power required can be easily provided by one transistor inverter; therefore, the $B+$ supply can be a single inverter with multiple outputs and regulators.

With these in mind, the over-all system approach can be discussed.

Fig. 1 shows a block diagram of the system described in the following discussion. The $B+$ supply consists of a silicon transistor inverter with a number of output voltages. Each of these voltages is fed to a magnetic amplifier where it is pulse-width modulated. Next, the magnetic amplifier output is filtered to produce the required direct voltage. A portion of the output power is fed back to a control winding on the magnetic amplifier and compared magnetically with a stable reference. The difference between these two signals will control the pulse width of the unfiltered direct voltage to keep the output voltage constant. In order to reduce the gain requirements of the pulse-width modulator, part of the regulation is accomplished by open-loop compensation sensitive to the input voltage. The output voltages can be individually adjusted by potentiometers in the feedback loops. The complete $B+$ supply consists of five magnetic amplifier regulators, each fed from a separate secondary winding on the inverter output transformer. The remaining two outputs, +230-volts and +300-volts 15-ma load, are not individually regulated but are produced by combinations of other outputs. To help minimize the number of components required, a single reference is provided for all the magnetic amplifiers in the $B+$ supply. Also, since the frequency of operation and the sequence of starting allow the $B+$ inverter to be driven from the filament supply, this method is adopted. Not only does this

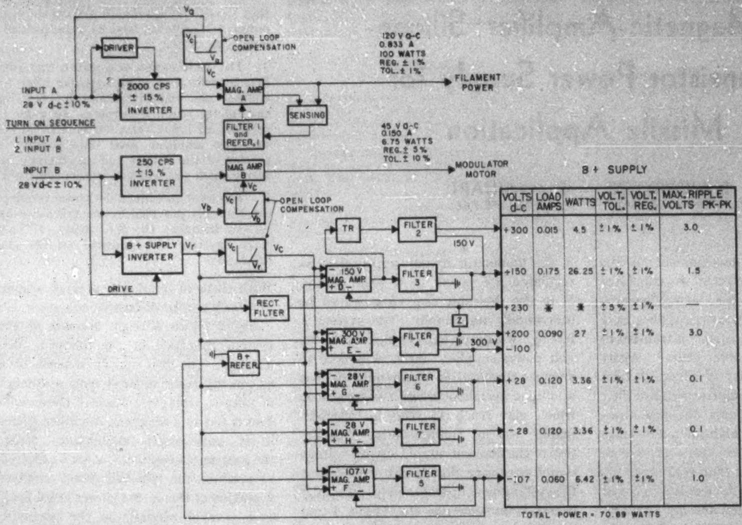
---

Paper 59-1054, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-IRE Nonlinear Magnetics and Magnetic Amplifiers Conference, Washington, D. C., September 23–25, 1959. Manuscript submitted June 22, 1959; made available for printing July 17, 1959.

B. Mokrytzki and R. A. Stuart are with Westinghouse Electric Corporation, Pittsburgh, Pa.

Dr. Werner Leonhard, formerly of Westinghouse Electric Corporation, contributed significantly to the development of this power supply.

Fig. 1. Over-all block diagram


INPUT A
28 V d-c ± 10%
TURN ON SEQUENCE
1. INPUT A
2. INPUT B
DRIVER
V_a
V_c
OPEN LOOP COMPENSATION
2000 CPS ± 15% INVERTER
MAG. AMP A
SENSING
FILTER 1 and REFER. 1
120 V a-c
0.833 A
100 WATTS
REG. ± 1%
TOL. ± 1%
FILAMENT POWER
INPUT B
28 V d-c ± 10%
250 CPS ± 15% INVERTER
MAG. AMP B
45 V a-c
0.150 A
6.75 WATTS
REG. ± 5%
TOL. ± 10%
MODULATOR MOTOR
V_b
OPEN LOOP COMPENSATION
B + SUPPLY INVERTER
V_r
DRIVE
TR
FILTER 2
150 V
− 150 V MAG. AMP + D −
FILTER 3
RECT FILTER
Z
− 300 V MAG. AMP + E −
FILTER 4
300 V
B+ REFER
− 28 V MAG. AMP + G −
FILTER 6
− 28 V MAG. AMP + H −
FILTER 7
− 107 V MAG. AMP + F −
FILTER 5

B + SUPPLY

VOLTS d-c	LOAD AMPS	WATTS	VOLT. TOL.	VOLT. REG.	MAX. RIPPLE VOLTS PK-PK
+300	0.015	4.5	±1%	±1%	3.0
+150	0.175	26.25	±1%	±1%	1.5
+230	*	*	±5%	±1%	—
+200 −100	0.090	27	±1%	±1%	3.0
+28	0.120	3.36	±1%	±1%	0.1
−28	0.120	3.36	±1%	±1%	0.1
−107	0.060	6.42	±1%	±1%	1.0

TOTAL POWER = 70.89 WATTS

* NEGLIGIBLE

reduce the number of components in the $B+$ inverter, but it also precludes any starting difficulty which might be encountered with a self-excited circuit.

For the filament supply, a low-power silicon transistor oscillator supplies power to drive a high-power silicon transistor inverter. The square-wave output of the power inverter is pulse-width modulated by a magnetic amplifier to produce the filament output voltage. This output voltage is metered by an rms sensing circuit and compared to a Zener reference in a voltage-error detector. The resulting error signal is fed to a control winding on the magnetic amplifier which then corrects the output. Open-loop compensation is again used to reduce the feedback requirements. In order to withstand the 45-volt overshoot referred to previously, high-voltage transistors are used throughout the filament supply.

The modulator supply inverter is a self-excited inverter circuit in which the frequency of oscillation is determined by the output transformer saturation. This circuit is easy to start and, since it operates at a relatively low power level, the inefficiencies incurred by allowing the output transformer to saturate can be tolerated. The inverter output voltage is regulated by another magnetic amplifier controlled only by an open-loop signal sensitive to the direct-voltage input. No feedback loop is needed since the required voltage regulation is only ±5% of the nominal voltage.

Transistor Specifications

The transistors used in this supply are n-p-n fused junction silicon power transistors, Westinghouse type *WX 1015* and *WX 1016*. The voltage and current ratings available, together with a very low saturation resistance, enable them to handle high switching powers with a minimum of losses. They can be obtained in 2- and 5-ampere current ratings with collector-to-emitter voltage ratings from 30 to 300 volts. The maximum saturation resistance of the 5-ampere unit is 0.5 ohm, and the gain is 10 or greater at the rated current. From an efficiency standpoint, rough calculations showed that each 0.1-ohm unit of transistor saturation resistance decreased the over-all system efficiency by approximately 2%. To keep the efficiency as high as practical, a specification of $R \leq 0.3$ ohm was placed on the 5-ampere transistors used for this supply.

Magnetic Amplifier Design

Although, for the most part, conventional magnetic amplifier design techniques were applied to the synthesis of this unit, a considerable number of unique facets are evident. These are brought about by rather stringent requirements with respect to performance, reliability, size, efficiency, and environmental conditions.

Beyond the requirements already listed are a number of practical ideas which should be included in the design procedure. In essence, these are considerations which would permit standard production techniques to be applied to the finished unit. Included in these are reproducibility of the entire unit performance, availability of basic components such as cores, relative ease of winding, and relative ease of assembly.

In order to transmit an understanding of the problems peculiar to this application, several topics will be discussed in detail in the following section.

The filament and modulator motor supplies, being a-c outputs, require doubler magnetic amplifiers. In the d-c supply bridge magnetic amplifiers are used in all cases except the magnetic amplifier which produces the 150-volt output as a doubler. Bridge circuits are employed because of the high voltages encountered in some outputs; however, center-tap circuits could be used on the 28-volt outputs. The lone, $B+$, doubler is inserted in the d-c supply so that an additional 150-volt output can be generated through an isolation transformer and a rectifier then stacked on the directly formed 150-volt output to create a low-power 300-volt supply. A Zener diode string used in conjunction with a portion of the higher power 300-volt output creates the 230-volt supply.

Cores and Diodes

Assuming that the performance requirements are all met, the prime magnetic amplifier design factor is size. Since the amplifiers will occupy a significant part of the available space, their volume must be minimized. A critical step in satisfying these conditions is the choice of a core material. Two metals which prove suitable for low-power amplifiers are HyMu-80 and Hipernik V. Of the two, HyMu-80 in tape-wound toroidal form is chosen for a number of reasons. However, in this case, the chief asset of HyMu-80 is its ability to retain better its gain at higher frequencies. Although this material has a lower maximum flux density than does Hipernik V, the fact that the amplifier need exercise only partial control of the input due to the relatively small input variations, coupled with the fact that a high-frequency supply does exist, reduces the load-winding requirements to a minimum. It was found by quick trial design that the smallest standard core of either material which could possibly fulfill the volt-second requirements of each output and still allow a reasonable control-winding area was of such a size as to require only a relatively small fraction of its area for the load winding. Later, it will be shown in the case of the d-c supply that a large amount of control-winding space will tend to reduce the size of the reference while also having a beneficial effect on efficiency. Thus, if the majority of the winding area is devoted to control turns, the maximum flux density of a core has a less important effect on its final dimensions.

Results of 2-kc tests which involved HyMu-80 in tape thickness of 1 and 2 mils displayed that the 1-mil tape did not possess any significant advantage over the heavier strip; therefore, 2-mil tape was used throughout.

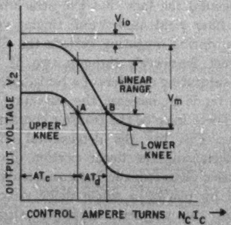


Fig. 2. Magnetic amplifier control characteristics

The filament and modulator motor supplies, utilizing doubler circuits, call for self-saturating diodes of relatively low voltage ratings and not too strict leakage values. However, in the $B+$ supply, the bridge magnetic amplifiers require voltage ratings in excess of those dictated by the nominal value of the peak inverter outputs which are considerably higher than the output voltages since the full magnetic amplifier ranges are not utilized. This is due to the effect of stray capacities working with the filter inductance which generates high ringing voltages comparable to the peak inverter voltages. Utilizing *R-C* spike suppressors to damp out these oscillations, the ringing voltages are reduced and add only approximately 25% to the necessary voltage ratings of the diodes. In some cases, namely the 300-volt channel, two units were used in series. These diodes are chosen to possess extremely low leakage currents in order not to penalize the high gain of the cores at elevated temperatures.

Load and Control Windings

The two magnetic amplifier control characteristics of Fig. 2 display a major portion of the information needed for the entire design. The plots reveal the change of output voltage with respect to control current for the two cases of high and low extremes of input voltage. As the input voltage is raised from its minimum to its maximum value, the operating point on the characteristics will shift downward from *A* to *B*, thus maintaining the output at a constant value. It is apparent from these curves that in order to fully utilize the gain of each core, the operating points *A* and *B* must lie comfortably within the linear range. The necessary magnitude of this region is determined by the magnitude of the input voltage which, in turn, is dependent on: 1. the required output voltage, 2. the magnitude of the upper knee, 3. the saturation induced voltage V_{io}, 4. the internal voltage drops due to winding resistance, diodes, etc., and 5. temperature effects. The total voltage which the amplifier must be capable of supporting is then given by $V_M + V_{io}$ if internal voltage drops are neglected.

For a square-wave supply voltage,

$$V_M+V_{io}=4fN_LB_MA_C\times10^{-8} \qquad (1)$$

where

f = frequency of supply voltage, cycles per second (cps)
B_M = maximum flux density, gauss
A_C = core cross-sectional area, cm^2
N_L = number of load turns

It is important to note here that variations in core characteristics from unit to unit, or with temperature changes, cause V_M to be made large enough to insure operation in the linear range under all conditions.

In order to appreciate the demands of core nonlinearities on load-winding requirements, a comparison may be made between idealized and practical control characteristics. Assuming the absence of nonlinearities and saturated induced voltage, a magnetic amplifier must support a minimum of approximately 22% of the output voltage. In practice, however, a conservative design requires that the load windings be capable of supporting 50% of the output voltage. This corresponds to operation in the central 50% of the control characteristic.

It has been previously mentioned that the control windings occupy a major portion of the available winding space. In determining the design of these windings, it is desirable to obtain the highest circuit gain for the least amount of space. Since the output of each amplifier is to be separately monitored by a closed-loop feedback system, some consideration must be applied to the manner in which the reference and feedback signals are mixed. In the case of the filament supply amplifier, it is advantageous to utilize a Zener bridge reference as a part of an rms voltage-sensing unit. However, in the d-c supply, since several units operate in essentially the same manner, a common Zener voltage reference can be utilized, feeding virtually a constant current to a series connection made up of reference windings from each of the individual cores. Feedback ampere-turns which are derived from the filtered output of an amplifier are fed to a second of its windings and thereby are magnetically mixed with the reference. This method of ampere-turns comparison does away with providing a number of Zener bridge references which would prove to be bulky and would consume relatively large amounts of power on the high-voltage outputs.

For a closed-loop regulation of ±1% for a ±10% input change, a loop gain of approximately 10 is required. Transcribing the required gain to feedback ampere-turns, a 2% change in output must yield enough net ampere-turns variation to allow the magnetic amplifier to swing its entire control range. Thus, AT_d of Fig. 2 is equal to 2% of the required feedback ampere-turns. Note that AT_c is the ampere-turns necessary to drive the core into its linear range and is equal to from one to several times AT_d. Therefore, it can also be assumed that the reference ampere-turns AT_r are approximately equal to the feedback ampere-turns AT_f since $AT_f >> AT_d + AT_e$ and $AT_f - AT_r \leq AT_d + AT_c$, then

$$AT_f = \frac{AT_d}{0.02} \simeq AT_r \qquad (2)$$

The AT_d for a given core is obtainable from test data. Once the reference and feedback requirements are known, the control windings can be designed.

In fitting the load and control requirements which were previously attained to specific cores, it becomes evident that core gain has the most significant effect on the performance and size of an amplifier. It is found that the load windings are easily designed as per previous minimum requirements by providing a wire size which will yield a current density low enough to insure safe operation. On the other hand, the control design is complicated by the fact that the wire size is not firmly tied to the current which flows through these windings. For example, the feedback windings of necessity must draw a minimum of current from the output in order to retain a high over-all efficiency. This implies a high number of turns. If the choice of wire is only based on current density, the resulting wire size is too small to be wound on the core economically. Thus, a somewhat larger wire must be tolerated. This is one factor that tends to amplify the need for cores which provide high gain and sizable amounts of control-winding area. Another item which must be realized is that the reference current, whose maximum value is prescribed by the Zener diodes which are used in the reference circuit, and while determining the number of turns on the windings which it energizes, does not firmly indicate wire size. Power limitations on the reference will cause it to display a relatively low output impedance. Thus, in the reference circuit, the ratio of winding resistances to fixed resistance must be kept small. Or else, the resistance variations, due to the copper in the windings, will become so large as to defy suitable temperature compensation for constant current. Hence, larger wire sizes must again be utilized. All in all, the need is evident for high core gains to reduce the necessary feedback and reference turns while preserving efficiency and retaining a small size.

The description of the design to this point has provided the minimum loop gain which would meet the required performance. To alleviate the necessity of over-design of the feedback loops in order to anticipate variations of core characteristics from unit to unit or with temperature, open-loop compensation of each channel is provided. The effect of this addition is to reduce the output variations to approximately 25% of the values which would exist if the feedback were unaided.

It is important to note that in applying the windings to the cores, the high-frequency square-wave nature of the magnetic amplifier supply brings to light the problem of distributed winding capacities which affect core gains and yield spurious oscillation or ringing, and the high-control to load-turns ratios which cause large voltages to be induced in the control circuits, thus endangering the insulation of those windings. These effects are virtually eliminated by: 1. providing sufficient sectoring of the gate windings to minimize their distributed capacitances, and 2. adopting a winding procedure for the magnetic amplifiers which is carried out by first winding the load turns separately on to each core of a given pair, then stacking the two cores and winding all control windings over both cores simultaneously. Thus, the cores, if closely coupled, will experience practically no induced voltage in the control circuit, and, hence, reflected capacitance effects will also be eliminated.

REFERENCE

The reference circuit itself was designed for economy with respect to efficiency and size. The fact that it is energized from one of the regulated outputs reduces its complexity, the necessary number of Zener diodes, the necessary magnitude of the input impedance and consequently the input power, while retaining a very stable output voltage.

One of the most serious problems of the entire $B+$ supply is an interaction of the magnetic amplifiers through the series-connected reference windings. This detriment is caused by the induced current from each magnetic amplifier which tends to flow in the reference winding loop and results in a degeneration of the gain of the individual magnetic amplifiers by a factor of 3 or 4. It was found that a 0.2-henry choke inserted in that loop tends to restore the gains to approximately one-half their isolated value. Further experimentation disclosed that encircling the 300-volt channel's reference winding and the series choke with a low-impedance capacitive path brought about a further reclamation of gains and thereby virtually eliminated the interaction problem. Since this last measure merely isolated the 300-volt channel from the others, the nucleus of the interaction problem can be attributed to that reference winding.

Although in the a-c supply a Zener bridge reference is used instead of ampere-turns comparison, the magnetic amplifier design procedure is essentially the same as the d-c procedure with respect to load and control windings.

The magnetic amplifier of the modulator motor supply differs in design from the others in the unit. Since the output is controlled entirely by open-loop compensation, core gain is of little consequence in this amplifier; however, linearity and a high maximum-flux density are desirable. Thus, Hipernik V becomes an excellent choice for this output. To further increase linearity while gaining uniformity of performance with varying core parameters, negative feedback is employed. This feedback is achieved by two resistors which shunt the self-saturating rectifiers. Their effect is also such as to bend the transfer characteristic about the lower knee, thus providing suitable biasing.

FILTERS AND STABILITY

The filter inductors for the $B+$ supply are wound on powdered iron cores. Again to reduce stray capacities, the windings contained many sectors. The filter capacitors for the high-voltage outputs are paper which proved in those cases to be more desirable than their electrolytic counterparts. However, on the 28-volt d-c outputs, tantalum units were employed. In general, the design for these elements is based on these factors:

1. Sufficient attenuation must be provided to reduce the output ripple to required amounts.
2. Continuous inductor currents must be insured in order to maintain coupling between the magnetic amplifiers and the loads
3. The size of the filters shall be minimized.

Beyond the fact that the attachment of a filter yields undesirable ringing voltages, during pregating periods the filter current freewheels through two of the self-saturating rectifiers of a bridge circuit and tends to increase the curvature at the lower knee of the transfer characteristic. This has the detrimental effect of decreasing the linear portion of the transfer characteristic but can be alleviated by placing a single diode directly across the magnetic amplifier output. Since the diodes which are used exhibit a virtually constant voltage characteristic, the effect of the freewheeling current is

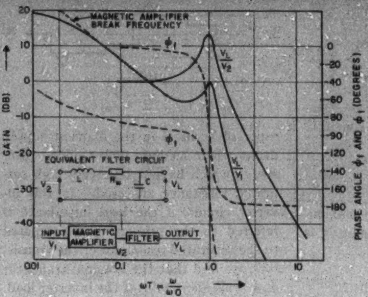


Fig. 3. Normalized Bode plots

reduced to one half its former value and the sharpness of the lower knee is markedly improved.

The design of the magnetic amplifier-filter combination must be such as to insure stability of each individual feedback system. The frequency characteristics of the magnetic amplifiers are accurately described by a single time-constant relationship since the response time is a large multiple of one cycle. The loop gain which is supplied by the amplifier has been specified by the allowable variations of the output. Thus the key to identifying the conditions for stability of the feedback system is the determination of the filter characteristics and the blending of these with the amplifier characteristics. Fig. 3 displays the normalized frequency plots for the gain V_L/V_2 and phase ϕ_f of a typical *L-C* filter. Note that a peaking of the response curve occurs at the filter break frequency and is a function of the damping factor *d*. The transfer function of the filter is derived using the equivalent circuit shown and is given in reference 1 as:

$$\frac{V_L}{V_2}=\frac{1}{\left(j\frac{\omega}{\omega_o}\right)^2+\left(j\frac{\omega}{\omega_o}\right)2d+1} \qquad (3)$$

$$\phi_f=\tan^{-1}\frac{2d\left(\frac{\omega}{\omega_o}\right)}{1-\left(\frac{\omega}{\omega_o}\right)^2} \qquad (4)$$

where

$$\omega_o=\frac{1}{\sqrt{LC}}\equiv\text{undamped natural frequency}$$

and

$$d=\frac{Rw}{2\sqrt{\frac{L}{C}}}\equiv\text{damping factor}$$

The damping effect due to the load is neglected since it is considerably smaller than that presented by *Rw* which is the filter resistance and effective magnetic amplifier output resistance.

Upon establishing the total response characteristic V_L/V_1, Fig. 3, it can be seen that the system is conditionally stable if the peak due to the filter characteristic is somewhat below the zero decibel or unity-gain line. Stability, therefore, is achieved by lag compensation, relying entirely on the single break given by the amplifier to reduce the loop gain to less than 1 before the high phase shift of the filter is expended. Thus the minimum allowable time constant of the magnetic amplifier is determined. This time constant has been shown from reference 2 to be:

$$T=\frac{G_t}{2f}\sum\frac{N_C^2}{R_C}\text{ seconds} \qquad (5)$$

where

N_C = control turns of a given winding
R_C = apparent resistance of that winding
G_t = Roberts' core gain $\left(\frac{\text{volts/turn}}{\text{ampere-turn}}\right)$

Since the feedback windings contribute essentially all the N^2/R the time constant of the magnetic amplifier is easily adjusted by means of a shunting resistor across this winding.

The amplifier time constant is made several times larger than the minimum values which are specified by the plots to allow for deviations of core and filter parameters. The shunting resistors have negligible effect on the feedback current itself.

RMS Sensing

The specifications given for the output voltage of the filament supply require regulation to within ±1% of its nominal rms voltage. For measuring rms values, multiplying devices such as Hall multipliers or square-law devices such as "starved diodes" or inherently rms sensitive elements such as thermocouples could be employed. None of these devices, however, provide a simple and uncritical solution over a wide operating temperature range with an accuracy of better than ±1%. A new type of rms sensing circuit using only small transformers and rectifiers can, however, meet these requirements. Figs. 4(A) and (B), which show a simplified block diagram of the filament system along with the associated voltage waveforms, will be used to explain the action of the rms sensing circuit.

The rms value of the output voltage (V_2) may be written as

$$V_{2_{rms}}=\sqrt{1-\epsilon}\,V_o \qquad (6)$$

where V_o is the peak amplitude of the filament output voltage and ϵ is the fractional part of a half period in which the magnetic amplifier absorbs the inverter voltage. The expansion of equation 6 into a series results in

$$V_{2rms}=V_o\left[1-\frac{\epsilon}{2}-\frac{1\cdot\epsilon^2}{2^2\cdot 2!}-\frac{1\cdot 3\cdot\epsilon^3}{2^3\cdot 3!}-\frac{1\cdot 3\cdot 5\cdot\epsilon^4}{2^4\cdot 4!}-\dots\right] \qquad (7)$$

If a linear approximation is made for $V_{2_{rms}}$ by using the first two terms only, then

$$V_{2rms}\approx V_o\left[1-\frac{\epsilon}{2}\right] \qquad (8)$$

By comparing this first approximation with the exact series representation of $V_{2_{rms}}$, it is obvious that a better approximation can be made; notice that the

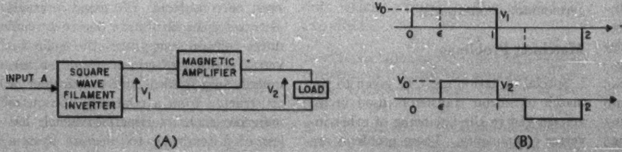
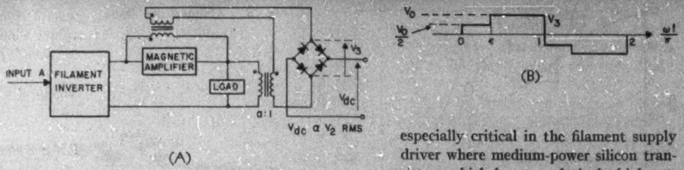


Fig. 4. Filament supply
A—Block diagram
B—Waveforms

Fig. 5. Rms circuit (A) and waveforms (B)



error is zero for $\epsilon = 0$ and that the value of the linear approximation suggested is always greater than the exact function for any $\epsilon > 0$. If the approximation is changed so that

$$V_{2_{rms}} \approx V_o[1.003 - 0.56\epsilon] \tag{9}$$

the error will be reduced for larger values of ϵ, i.e., $\epsilon > 0.25$. The maximum ϵ necessary to regulate the output voltage of the filament supply, calculated from input voltage variations and including a small safety margin to account for internal losses, is 0.4. The error over the range of $0 < \epsilon < 0.4$ for the approximation in equation 9 is +0.35%, −0.64% and is compatible with system requirements. Fig. 5(A) shows how the linear approximation can be realized with only two small transformers and a bridge rectifier. The input voltage to the rectifier bridge has the waveform in Fig. 5(B). Its rectified average value corresponds to the expression in equation 9 and is proportional to the rms output voltage.

In the preceding derivation, it was assumed that the filament supply had a zero output impedance. In a practical circuit the peak voltage seen by the rms sensing circuit during the time $0 < t < \epsilon T/2$, when the load draws no current, is higher than the voltage seen between $\epsilon T/2 < t < T/2$ when the load is receiving power. This necessitates the application of a correction factor b, representing the regulation of the filament supply from no load to full load, to equation 9 so it now becomes

$$V_{2_{rms}} \approx V_o\left[1.003 - \frac{0.56\epsilon}{b}\right] \tag{10}$$

and b is defined as $b = V_o'/V_o$, where V_o' is the peak inverter output voltage during time $0 < t < \epsilon T/2$. The evaluation of b is done by experiment or can be calculated since it depends only upon the internal impedance and the load impedance of the filament supply and is not a function of ϵ. The usefulness of this type of rms circuit is restricted by regulation to constant load circuits; i.e., for an application where the load impedance varies, b would vary and introduce a sensing error. The error can be minimized by reducing the internal impedance of the supply as much as is practical.

Starting Conditions

The characteristics of the filament load impose difficult starting conditions on the filament inverter since the initial impedance of the cold filaments is approximately one fourth its normal value. If the filament inverter is self-excited, the transistors must be provided with sufficient drive to remain in saturation during the starting condition in order that oscillations be maintained. The silicon transistors used have a gain of 10 at 5 amperes collector current; at 20 amperes collector current, the gain would be down to 5 or less so that the required drive would be at least eight times normal. A considerable amount of power would be needed to provide the necessary excess drive as dictated by the starting surge. The solution chosen is to use a small oscillator to provide drive to the main inverter and to let the main inverter power transistors limit the output current and voltage during starting by coming out of saturation. The main inverter transistors will limit the load current to some level determined by their drive currents and the output voltage will gradually build up to its normal value as the transistors return to a saturated condition. The starting period requires from 5 to 7 seconds. During this transient, the power dissipation in each of the two transistors reaches a peak of 120 watts. In addition to the unusual load, the input voltage upon starting exhibits an overshoot to 45 volts for as long as 100 ms. The transistors must also be capable of sustaining this high input voltage. These abnormally severe transient conditions underline the superior ruggedness of silicon transistors when compared to their germanium counterparts.

Matching Problems

Special attention must be given to the match of silicon transistors used in inverters and to the balancing of other inverter components. These problems are especially critical in the filament supply driver where medium-power silicon transistors which have a relatively high saturated resistance are used. In this inverter circuit, if the load impedance is referred to the primary and compared to the saturated resistance of the transistors, we find that the nominal transistor drops represent 10% of the inverter load. If the drops of the two inverter transistors are not equal, a different voltage is applied to the timing transformer during opposite half cycles causing a time unbalance of the output voltage; i.e., the time interval for positive output voltage is not equal to the time interval for negative output voltage. When this unbalance is transmitted to the main inverter, the main transistors have unequal conduction times, introducing a direct-voltage component in the main inverter output transformer, causing the transformer to saturate for a time roughly equal to the difference in conduction angles, and resulting in current spikes, distortion of the output waveform, and abnormal losses. It is also possible to have a time unbalance in the driver output voltage due to an unbalanced driver load.

To solve these problems, two things are required. First, the saturated resistance of the driver transistors must be matched; second, the load on the driver must be balanced by matching the input impedances of the main power transistors. When these two parameters are matched, the output of the driver inverter will be a symmetrical square wave. This does not insure, however, that the main output transformer will not saturate from other conditions.

In the main inverter, the saturated resistance of its transistors represents about 5% of the load impedance so that here the problem is less severe. Also, the frequency of oscillation and conduction times have already been determined by the driver so that they are not affected by mismatched transistors in the main inverter; it is the output transformer that must be considered. To avoid intervals of transformer saturation due to a small direct-voltage component, the main inverter transistors are matched for saturated drop. Also, the transformer is constructed from a relatively nonsquare material, such as Hipernik, which has less of a tendency to saturate than a

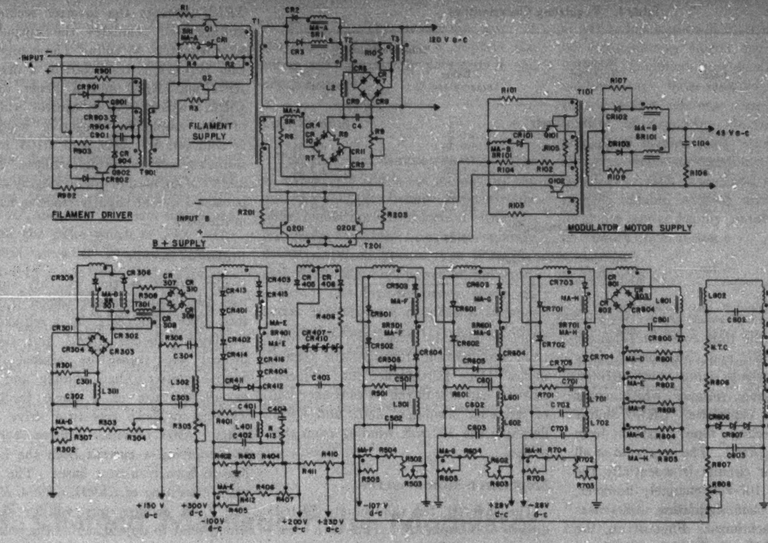

INPUT A
R1
SR1 MA-A
CR1
Q1
R4
R2
Q2
R3
T1
CR2
MA-A SR1
CR3
T2
T3
120 V a-c
R10
CR5
CR7
L2
CR9
CR8
FILAMENT SUPPLY
R901
CR901
Q901
CR903
R904
C901
R903
CR904
Q902
CR902
T901
R902
FILAMENT DRIVER
MA-A SR1
R6
CR4
CR10
R8
C4
R9
R7
CR11
CR6
INPUT B
R201
Q201
Q202
R202
T201
B + SUPPLY
R101
T101
CR101
Q101
MA-B SR101
R105
R104
R102
Q102
R103
MODULATOR MOTOR SUPPLY
R107
CR102
CR103
MA-B SR101
C104
R108
R106
43 V a-c
CR305
CR306
CR307
CR310
MA-D
R308
T301
CR308
CR309
CR301
CR302
CR304
CR303
R306
C304
R301
C301
L301
C302
L302
C303
R305
MA-D
R303
R307
R304
R302
+150 V d-c
+300V d-c
CR413
CR403
CR405
CR406
CR415
CR401
MA-E SR401
CR402
MA-E
CR414
CR416
CR404
CR411
CR412
R408
CR407– CR410
C403
C401
C404
R401
L401
C402
R413
R402
R403
R404
R410
R411
MA-E
R406
R412
R405
R407
-100V d-c
+200V d-c
+230V d-c
CR503
MA-F
CR501
SR501 MA-F
CR502
CR505
CR504
C501
R501
L501
C502
MA-F
R504
R502
R505
R503
-107 V d-c
CR603
MA-G
CR601
SR601 MA-G
CR602
CR605
CR604
C601
R601
L601
C602
L602
C603
MA-G
R604
R602
R605
R603
+28V d-c
CR703
MA-H
CR701
SR701 MA-H
CR702
CR705
CR704
C701
R701
L701
C702
L702
C703
MA-H
R704
R702
R705
R703
-28V d-c
CR801
CR802
CR803
CR804
L801
C801
CR805
MA-D
R801
MA-E
R802
MA-F
R803
MA-G
R804
MA-H
R805
L802
C802
MA-E
MA-D
N.T.C
MA-F
R806
MA-G
MA-H
CR806
CR807
CR808
C803
R807
R808

Fig. 6. System schematic

true square-loop material while retaining the low loss qualities of a narrow loop.

Temperature Compensation

A final, but significant, element in the development of the unit is the temperature compensation since the entire unit is subjected to temperatures ranging from −40 C to +100 C. In all amplifiers the variation of copper resistance in the load circuit is negligible. The d-c supply contains several temperature sensitive elements, the most important being the copper resistance of the reference loop which tends to lower the output by 3% with increasing temperature. Combatting this deviation by different amounts in different channels is the feedback winding resistance. Finally, the Zener diodes of the d-c reference provide a small fraction of a per-cent variation which could be positive or negative. The bulk of the compensation in the d-c supply is achieved by a negative temperature coefficient resistor in the reference loop while trimming in the form of positive temperature coefficient resistors in the feedback is required on the 107- and 150-volt magnetic amplifiers.

In the filament supply, any significant variation of the rms output voltage with temperature is the result of temperature-sensitive elements in the rms sensing circuit and the reference bridge since variations of AT_e or changes of copper resistance in the load circuit become negligible with a high loop gain. For the detector bridge, temperature-stable resistors are used so their variation can be neglected; however, the balance point can vary up to a maximum of ±0.6% due to the temperature variation of the Zener diode voltages. In the rms sensing circuit the winding resistances of the sensing transformers and filter choke will have a positive temperature coefficient, but with proper design this is counteracted by the negative temperature coefficient of the diode forward voltage drops in the rectifying bridge.

When subjected to temperature testing, the modulator magnetic amplifier output displayed a tendency to drift quite radically, enough to exhaust the allowable deviation. However, the gain or linearity held up quite well. Thus, the problem becomes one of compensation for the drift which is traced and attributed to two effects:

1. The variation of the rather high copper resistance of the magnetic amplifier and inverter transformer causes changes in the output voltage.
2. The variation in the magnetic amplifier winding resistance causes an output deviation through the feedback loop, since the amount of negative feedback is proportional to the load current and the winding resistance.

Both of these variations are negative temperature effects when referred to the output; hence by using a higher voltage Zener diode, whose temperature coefficient is higher (and positive), compensation is achieved.

Results

B+ Supply

Fig. 6 displays the complete schematic diagram of the final system. Taking the 107-volt amplifier *SR501* as an example, it can be seen that its output is fed to a filter composed of *L501* and *C502* resulting in a smooth d-c output for this channel. This voltage is then fed through *R502*, *503*, and *504* to a feedback winding of *MA-F*. The reference voltage is formed across the Zener diode combination, *CR806*, *807*, and *808*. The voltage level of the output can be adjusted by potentiometer *R503* which varies the net ampere-turns on each core. *R504* is a temperature-sensitive resistor. Open-loop compensation is accomplished by the network containing *CR805* which furnishes a current through *R803* to an amplifier winding. *R505* is used to increase the N^2/R of the amplifier and there-

Table I. Regulating Characteristics

Output	Load, Amperes	Magnetic Amplifier Number	Total Deviation: Room Temperature, %	Total Deviation: −40 C to +85 C, %
		D-C Supply		
+300	0.015		±0.2	±0.39
+300	0.090	SR401	±0.38	±0.54
+230	Negligible		±0.46	±0.74
+150	0.175	SR301	±0.25	±0.33
−107	0.060	SR501	±0.115	±0.15
+28	0.120	SR601	±0.05	±0.13
−28	0.120	SR701	±0.14	±0.24
		A-C Filament Supply		
120	0.833	SR1	±0.37	±0.51
		Modulator Motor Supply		
45	0.150	SR101	±0.17	±2.8

by increase the time constant to insure stability. *R501* and *C501* form a spike suppressor which damps out high-frequency ringing voltages which occur due to stray capacities. *CR505* shunts the freewheeling filter current from *CR501* and *CR502*. The operation of all the regulator circuits is essentially the same as the 107-volt channel; however, some small, but significant, differences should be mentioned. First of all, the two identical 28-volt channels, *SR601* and *SR701*, contain two stages of filtering which are necessary to obtain the low ripple content required in these inputs. The voltages encountered in the 300-volt channel necessitate two diodes in series. An additional spike suppressor formed by *C404* and *R413* is used to further alleviate the voltage stress on these rectifiers. The 230-volt output is also derived from the 300-volt channel by the resistive divider, *R402*, *R403*, and *R404*, along with the Zener diode supply which contains *CR407* to *CR410*. *R410* controls the output level of this supply. *R408* and *C403* form an *R-C* filter which is tolerable on this channel since the ripple content of the input is extremely low. The network containing *CR805* accomplishes open-loop compensation for all the amplifiers.

Several factors involved in the reference circuit are worthy of note. The network is fed by the 107-volt supply which is virtually constant. *R808* controls the current through the Zener diodes. Due to the fact that induced currents can flow in the series connection of reference windings, *L802* and *C802* decouple the 300-volt winding from the others.

MODULATOR MOTOR SUPPLY

The inverter for this supply is fed directly off input *B*. This is a conventional self-excited inverter circuit which produces a 250-cycle square wave. The core is a Hipernik V toroid and the transistors are Westinghouse *WX 1015C* types. *R101* and *R103* determine the base drive while *R105* is used to insure starting. A doubler magnetic amplifier, *SR101*, accepts the inverter secondary voltage and regulates the output to approximately 45 volts. The compensation network is composed of a divider, *R102* and *R104* and a Zener diode, *CR101*, which feeds a control winding of the amplifier. Negative feedback is achieved by *R107* and *R108* which shunt the self-saturating rectifiers. *R106* and *C104* are used to inhibit ringing which is caused by the slightly reactive nature of the load.

FILAMENT SUPPLY

The driver portion of this supply uses two medium-power silicon transistors, *Q901* and *Q902*, in a self-excited oscillator. Diodes *CR901* and *CR902* protect the transistors against high reverse emitter-to-base voltage, and also permit the use of a single feedback winding for both transistor drive currents. *R903* provides positive starting by supplying a current from the input line to both transistor bases. The network consisting of *CR903*, *CR904*, *R904*, and *C901* reduces any voltage spikes which tend to appear across the transistors. The driver supplies base current to two high-power silicon transistors (*WX 1016C*), *Q1* and *Q2*, in the filament supply inverter. The secondary voltage of *T1*, the filament supply output transformer, is pulse-width modulated by *SR1*, a magnetic amplifier, to produce the reg-

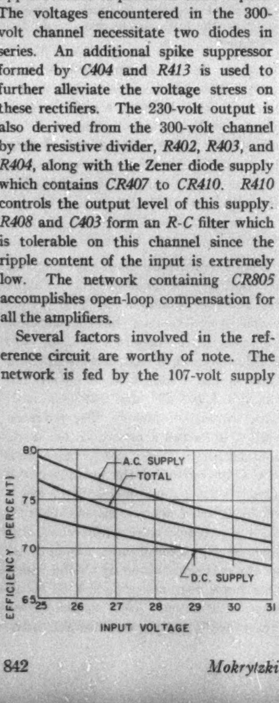


Fig. 7 (left). Efficiencies

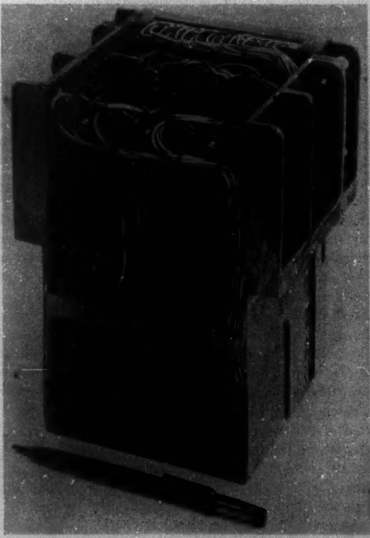


Fig. 8 (right). Completed prototype

ulated output voltage. Transformers *T2* and *T3*, along with a diode bridge and a filtering network, *L2* and *C4* make up the rms sensing circuit and produce a direct voltage proportional to the rms of the filament output alternating voltage. This direct voltage is compared to a Zener diode reference and an error signal produced in the voltage error detector consisting of *CR4*, *CR5*, *CR10*, *CR11*, *R7*, and *R8*. The error signal then controls the pulse width of *SR1*. The magnetic amplifier is open-loop compensated by *R2*, *R4*, and *CR1*. Two low-voltage secondaries on *T1* supply the drive power for the *B*+ inverter.

The results of tests which were performed on the completed unit are shown in Table I. Plots of efficiency versus voltage for the filament supply, the *B*+ supply, and the complete unit are shown in Fig. 7. The maximum over-all efficiency is 77%. Fig. 8 displays the entire unit successfully packaged in the allotted space.

Conclusions

This paper has demonstrated a number of advantages which are realized by the use of multiple-channel magnetic-amplifier regulators in conjunction with high-frequency inverters using the recently developed Westinghouse type *WX 1016* and *WX 1015* high-power silicon transistors. The high-frequency static inverter conveniently supplies a number of output voltages which can be regulated by compact, efficient magnetic amplifiers. The silicon transistors add to the inverter a ruggedness and flexibility previously unattainable in germanium units as well as low losses which are unmatched by contemporary silicon devices. The magnetic amplifiers provide reliable regulators which have isolated inputs so that a single reference can control a number of regulating channels. This alloying of magnetic and semiconductor elements yields an amplification of the qualities of each device and produces a rugged, reliable, efficient and high-performance power supply suitable for use in military applications.

References

1. Servomechanisms and Regulating System Design, Vol. I (book), Harold Chestnut, Robert W. Mayer. John Wiley & Sons, Inc., New York, N. Y., 1951.

2. Magnetic Characteristics Pertinent to the Operation of Cores in Self-Saturating Magnetic Amplifiers, R. W. Roberts. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 73, 1954 (Jan. 1955 section), pp. 682–90.

D-C Charged Magnetic Pulse Modulator

E. W. MANTEUFFEL ASSOCIATE MEMBER AIEE

R. E. COOPER ASSOCIATE MEMBER AIEE

AN A-C CHARGED magnetic pulse modulator consisting of a resonant charging stage and several successive pulse-modulating stages is the subject of a paper by W. S. Melville.[1] These pulse-modulating stages employ saturable reactors used as switching elements, and a series of capacitors. This modulator, if designed for a high-pulse repetition frequency, requires a separate frequency converting device whose output frequency must be equal to the desired pulse repetition frequency. Either an alternator or a magnetic frequency multiplier can be employed for this purpose. However, the magnetic frequency multiplier has considerable weight and volume, as well as a low power factor for high ratios of frequency conversion. An alternator, if built for a low speed, will also have a large weight. If designed for high speed, it has a limited life resulting from problems with bearings. To have a jitter-free-operation of the radar system, the voltage and frequency of the supplying power source must be regulated to close tolerances.

Melville also discusses d-c charged modulators employing thyratron tubes as discharges switches. Like many other electronic tubes, thyratons have the disadvantage of being less reliable than comparable solid-state devices. The recently developed controlled silicon rectifier, an essentially solid-state thyratron, opens new possibilities for further development of d-c charged magnetic pulse generators suitable for radar systems.

This paper presents information about investigations on a proposed d-c charged magnetic pulse modulator which partially employs controlled rectifiers as switching elements. A breadboard modulator was constructed which consisted of one charging and two pulse-modulating stages. One controlled rectifier was used in the first stage, two in parallel in the second stage, and a saturable reactor in the output stage. The modulator was capable of delivering a peak power output of approximately 110 kw (kilowatts) into a resistive load under critical damping. When a 5-section pulse-forming network was employed in the output stage, a 50-kw peak-power output pulse of about 3-μsec (microsecond) duration was achieved.

It should be emphasized that the work reported here is only a first step and that experiments were made only with a model-

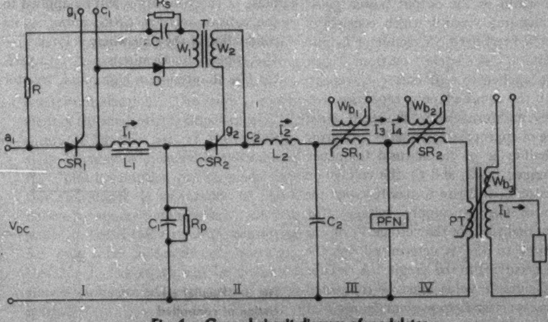


Fig. 1. General circuit diagram of modulator

Paper 59-1051, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-IRE Nonlinear Magnetics and Magnetic Amplifiers Conference, Washington, D. C., September 23–25, 1959. Manuscript submitted June 16, 1959; made available for printing August 4, 1959.

E. W. Manteuffel and R. E. Cooper are with the General Electric Company Advanced Electronics Center at Cornell University, Ithaca, N. Y.

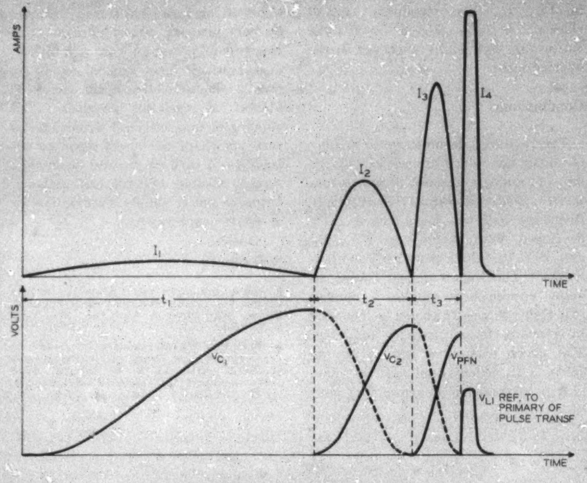


Fig. 2. Current and voltage wave shapes

type modulator. With further developments of controlled rectifiers of larger current ratings, much higher peak output powers than those obtained during this investigation will be possible. No special effort was made to reduce the pulse width to fractions of a microsecond, which would further increase the output power, since means for achieving this goal are described in Melville's paper and other publications.[1,2]

Description of Operation

The basic circuit configuration of the proposed modulator, which consists of several cascaded stages, I, II, III, IV, is illustrated in Fig. 1. The number of stages necessary depends on the required pulse width of the output power pulse. The charging stage 1 which consists of controlled rectifier CSR_1, inductor L_1, and capacitor C_1, is switched on from trigger pulses applied to terminals g_1-c_1, repeating at the required pulse-repetition frequency, and furnished from a small power source of appropriate frequency. During one half-cycle of the natural frequency determined by L_1 and C_1, the voltage on capacitor C_1 will rise to nearly twice the value of the d-c supply voltage V_{DC}. The actual ratio of the voltage V_1 on the capacitor to V_{DC} is determined by the losses occurring in the circuit. A resistor R_p of suitable value must be connected across capacitor C_1 to prevent charging of this capacitor through the anode-cathode resistance of CSR_1 to the supply voltage V_{DC} before trigger pulses are applied.

At the instant when V_1 reaches its maximum value, current I_1 will become zero and the controlled rectifier CSR_1 will return to its nonconducting state. This means that the voltage $V_1 - V_{DC}$ will suddenly appear between the cathode and anode terminals c_1 and a_1 of this rectifier. This sudden voltage step is now applied to a circuit consisting of a small capacitor C, resistor R, diode D, and the primary winding W_1 of a small pulse transformer T. The secondary winding W_2 of T is connected to gate terminal g_2 and cathode terminal c_2 of a second controlled rectifier CSR_2, located within stage II of the modulator. This second stage consists of inductor L_2 and capacitor C_2. A current pulse of short duration will be applied to the gate of rectifier CSR_2, initiating its conduction at the time where voltage $V_1 - V_{DC}$ appears at terminals a_1-c_1. Capacitor C is shunted by a high-ohmic resistor R_3 which allows its discharge between pulses.

After the time $t_2 = \pi\sqrt{C_0 L_2}$, (made by proper choice of the value of L_2 much shorter than the time $t_1 = \pi\sqrt{C_1 L_1}$) the charge of capacitor C_1 will be transferred to capacitor C_2. A current pulse I_2 of large magnitude and short duration will flow within stage II. The value C_0 is determined by

$$C_0 = \frac{C_1 C_2}{C_1 + C_2}$$

At the end of the time interval t_2, the voltage V_2, because of the influence of losses in stage II, becomes somewhat smaller than voltage V_1, and will exist at the terminals of capacitor C_2. At the same time, flow of current I_2 in CSR_2 will cease and this rectifier will return to the nonconducting state.

Shortly after voltage V_2 has attained its maximum value, saturable reactor SR_1 of stage III, whose unsaturated impedance is designed to be many times larger than the apparent impedance of L_2, will become saturated. Transfer of the charge from C_2 into the capacity C_p of the pulse-forming network PFN will take place by way of the saturated inductance L_{s_1} of the saturable reactor SR_1. L_{s_1} is made much smaller than L_2 to achieve further compression of the pulse width. Transfer of pulse power to the load by saturable reactor SR_2 and pulse transformer PT is self-explanatory. Saturable reactors SR_1 and SR_2 and the pulse transformer are furnished with bias windings W_{b_1}, W_{b_2}, and W_{b_3}. These serve to reset the flux during interpulse periods. The cores have a relatively long time available for resetting between pulses; therefore, the bias current and number of bias turns can be small.

The sequence of pulse currents and voltages as they will appear in the described modulator are illustrated in Fig. 2. Actual time compression ratios and differences in magnitude of pulse currents are much larger than shown in this qualitative illustration.

The output power rating of the modula-

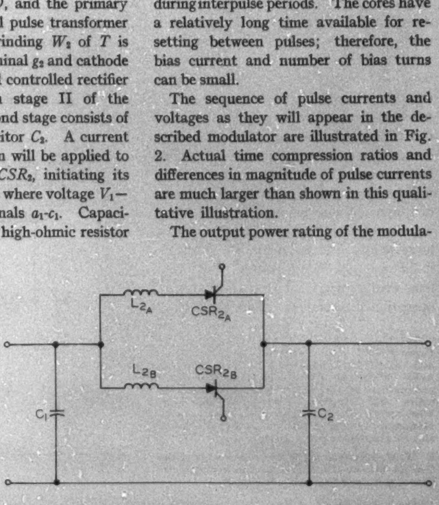


Fig. 3. Parallel operation of controlled rectifiers

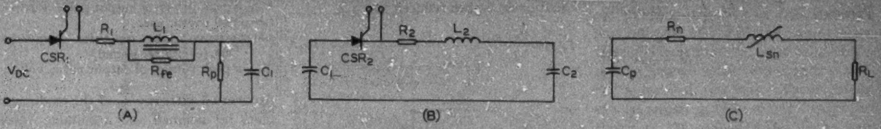


Fig. 4. Circuit diagrams of the various stages

A—First stage
B—Intermediate
C—Output stage

tor can be increased considerably by the use of two or more controlled silicon rectifiers in parallel in the second stage. This is shown in Fig. 3 where two controlled rectifiers, CSR_{2A} and CSR_{2B}, each in series with an inductance L_{2A}, L_{2B}, are arranged in parallel connection. The currents in both rectifiers will be essentially equal if they are triggered simultaneously.

The sequence of events will now be analyzed:

1. *Charging Stage.* The circuit of the charging stage is shown in more detail in Fig. 4(A). The inductance L_1 is shunted by a resistance R_{fe} representing the core losses. Resistor R_1 includes winding resistance and eddy current losses within the winding as well as dissipation in capacitor C_1. The purpose of resistor R_p was mentioned earlier.

Using Laplace notation the charging current I_1 is expressed by

$$I_1(s)=\frac{V_{DC}}{s\left[\dfrac{R_{fe}s}{s+\dfrac{R_{fe}}{L_1}}+\dfrac{1}{C_1}\left(\dfrac{1}{s+\dfrac{1}{R_pC_1}}\right)+R_1\right]} \qquad (1)$$

After rearranging terms and taking into account that R_p and $R_{fe}>>R_1$

$$I_1(s)=\frac{V_{DC}}{R_{fe}}\times \frac{s^2+\left(\dfrac{R_{fe}}{L_1}+\dfrac{1}{R_pC_1}\right)s+\dfrac{R_{fe}}{R_pC_1L_1}}{s\left[s^2+\left(\dfrac{R_p+R_{fe}}{R_pR_{fe}}\dfrac{1}{C_1}+\dfrac{R_1}{L_1}\right)s+\dfrac{1}{L_1C_1}\right]} \qquad (2)$$

This equation is of the form

$$I_1(s)=\frac{V_{DC}}{R_{fe}}\frac{s^2+a_1s+a_0}{s[(s+\alpha)^2+\beta^2]}$$

The solution in the time domain can be found in any table of Laplace transforms. Since the frequency of the damped oscillation will be about equal to the undamped natural frequency $1/\sqrt{C_1L_1}$, the charging current is given as a function of time with sufficient accuracy by

$$I_1(t)=V_{DC}\left[\frac{1}{R_{fe}}+\sqrt{\frac{C_1}{L_1}}\,e^{-\alpha t}\sin\frac{1}{\sqrt{C_1L_1}}\,t\right] \qquad (3)$$

where

$$\alpha=\frac{1}{2}\left[\frac{R_1}{L_1}+\frac{R_p+R_{fe}}{R_pR_{fe}}\left(\frac{1}{C_1}\right)\right] \qquad (4)$$

The voltage across C_1 is expressed as

$$V_{c1}(s)=I(s)\left(\frac{1}{C_1}\right)\frac{1}{s+\dfrac{1}{R_pC_1}} \qquad (5)$$

By combining equations 1 and 5 one obtains

$$V_{c1}(s)=\frac{V_{DC}}{R_{fe}C_1}\times \frac{s+\dfrac{R_{fe}}{L_1}}{s\left[s^2+\left(\dfrac{R_p+R_{fe}}{R_pR_{fe}}\left(\dfrac{1}{C_1}\right)+\dfrac{R_1}{L_1}\right)s+\dfrac{1}{L_1C_1}\right]} \qquad (6)$$

which is of the general form

$$V_{c1}(s)=\frac{V_{DC}}{R_{fe}C_1}\frac{s+a_0}{s[(s+\alpha)^2+\beta^2]}$$

For all practical purposes the voltage V_{c1} as a function of time is given by

$$V_{c1}(t)=V_{DC}\left(1-e^{-\alpha t}\cos\frac{1}{\sqrt{C_1L_1}}\,t\right) \qquad (7)$$

where α is again determined by equation 4.

2. *Intermediate Stage.* The circuit of any intermediate stage is illustrated separately in Fig. 4(B). Notations used are for stage II. Inductor L_2 is not shunted by a resistor in this case, since a pure air-core inductance is assumed. Resistor R_2 represents winding resistance including skin and proximity effects as well as capacitor dissipation. The series equivalent of the magnetizing impedance of the saturable reactor SR_1 of the following stage III may be included in the value of R_2. This assumption can be made since this impedance is essentially of ohmic nature for square-loop reactors. Current and voltages of this circuit are considered for only one half-cycle of operation. Resistor R_p across C_1 can be neglected, since its influence on the transfer of charge is small.

The transient current I_2 of this discharge circuit is given in Laplace notation by

$$I_2(s)=\frac{V_{c1\,\max}}{s\left(L_2s+R_2+\dfrac{1}{C_0s}\right)} \qquad (8)$$

where

$$C_0=\frac{C_1C_2}{C_1+C_2}$$

and $V_{c_1\,\max}$ is equal to the voltage attained at capacitor C_1 at the instant when current I_1 in stage I ceases to flow. After rearranging terms

$$I_2(s)=\frac{V_{c1\,\max}}{L_2\left(s^2+\dfrac{R_2}{L_2}s+\dfrac{1}{C_0L_2}\right)} \qquad (9)$$

or

$$I_2(t)=\frac{V_{c1\,\max}}{L_2\sqrt{\dfrac{1}{C_0L_2}-\left(\dfrac{R_2}{2L_2}\right)^2}}\,e^{-(R_2/2L_2)t}\times \sin\sqrt{\frac{1}{C_0L_2}-\left(\frac{R_2}{2L_2}\right)^2}\,t \qquad (10)$$

Since for all practical purposes the term $(R_2/2L_2)^2$ can be neglected

$$I_2(t)\cong V_{c1\,\max}\sqrt{\frac{C_0}{L_2}}\,e^{-(R_2/2L_2)t}\times \sin\frac{1}{\sqrt{C_0L_2}}\,t \qquad (11)$$

The voltage across capacitor C_2 is defined by

$$V_{c2}(s)=\frac{V_{c1\,\max}}{C_2L_2}\frac{1}{s\left(s^2+\dfrac{R_2}{L_2}s+\dfrac{1}{C_0L_2}\right)}$$

$$V_{c2}(t)\cong\frac{V_{c1\,\max}}{C_2L_2}C_0L_2\left[1-e^{-(R_2/2L_2)t}\times \cos\frac{1}{\sqrt{C_0L_2}}\,t\right] \qquad (12)$$

$$V_{c2}(t)\cong V_{c1\,\max}\frac{C_1}{C_1+C_2}\times \left[1-e^{-(R_2/2L_2)t}\cos\frac{1}{\sqrt{C_0L_2}}\,t\right] \qquad (13)$$

The voltage at capacitor C_1 as a function of time is then determined by

$$V_{C_1}(t) \cong V_{C_1\,\max}\left[1-\frac{C_2}{C_1+C_2}\left(1-\epsilon^{-(R_2/2L_2)t}\times \cos\frac{1}{\sqrt{C_0L_2}}t\right)\right]$$

$$\cong V_{C_1\,\max}\frac{C_2}{C_1+C_2}\left[\frac{C_1}{C_2}+\epsilon^{-(R_2/2L_2)t}\times \cos\frac{1}{\sqrt{C_0L_2}}t\right] \quad (14)$$

The latter equation indicates whether or not complete transfer of charge from C_1 to C_2 during the time of one half-cycle $t_2=\pi\sqrt{C_0L_2}$ has been achieved. This means that the condition

$$C_2=C_1\epsilon^{\pi/2(R_2\sqrt{C_0/L_2})} \quad (15)$$

must be fulfilled for complete discharge of C_1.

3. *Output Stage.* Fig. 4(C) illustrates the circuit of the output stage in a simplified form. The load R_L is referred to the input terminals of the output pulse transformer *PT*; see also Fig. 1. Effects of eddy current losses within the windings of reactor SR_n, the pulse transformer, and dissipation in capacitor C_p are included in the value of resistor R_n. L_{sn} represents the saturated inductance of the final saturable reactor SR_n. Leakage inductance of the pulse transformer should be included in this value.

A pulse-forming network which can shape the output pulse into a fairly rectangular form will replace capacitor C_p and inductance L_{sn} where the latter will be a component of this network. The network will be charged from the preceding stage nearly like a pure capacitor since its distributed inductance is small when compared to the saturated inductance of the reactor located within this stage. Details on design of pulse forming networks are given in reference 3. The simple discharge of C_p through L_s into the load resistor R_L will be discussed here.

The final discharge current I_n is expressed in Laplace notation by

$$I_n(s)=\frac{V_{cp\,\max}/L_{sn}}{s^2+\dfrac{R_L+R_n}{L_{sn}}s+\dfrac{1}{C_pL_{sn}}} \quad (16)$$

$V_{cp/\max}$ is the voltage attained at capacitor C_p at the instant when flux reversal in the previous saturable reactor takes place and current essentially ceases to flow in the preceding stage.

If the discharge takes place in a single unidirectional pulse, the maximum peak power will be obtained for critical damping; that means if

$$R_L+R_n=2\sqrt{L_{sn}/C_p}$$

equation 16 is then reduced to

$$I_n(s)=\frac{V_{cp\,\max}}{L_{sn}}\frac{1}{\left(s+\dfrac{1}{\sqrt{L_{sn}C_p}}\right)^2} \quad (17)$$

or

$$I_n(t)=\frac{V_{cp\,\max}}{L_{sn}}\,t\,\epsilon^{-t/\sqrt{L_{sn}C_p}} \quad (18)$$

The peak current and power will occur at time $t_p=\sqrt{L_{sn}C_p}$ and will have the values

$$I_{n\,\max}=V_{cp\,\max}\sqrt{\frac{C_p}{L_{sn}}}\,\frac{1}{\epsilon} \quad (19)$$

$$P_{\max}=\left(\frac{V_{cp\,\max}}{\epsilon}\right)^2\frac{C_p}{L_{sn}}R_L \quad (20)$$

When underdamped discharge occurs, the maximum obtainable peak power will be furnished to the load during the first half-cycle of discharge if $R_L+R_n=\sqrt{L_{sn}/C_p}$. Under this condition

$$I_n(s)=\frac{V_{cp\,\max}}{L_{sn}}\frac{1}{s^2+\dfrac{1}{\sqrt{L_{sn}C_p}}s+\dfrac{1}{L_{sn}C_p}} \quad (21)$$

or

$$I_n(t)=V_{cp\,\max}\sqrt{\frac{4C_p}{3L_{sn}}}\,\epsilon^{-(1/\sqrt{4L_{sn}C_p})t}\times \sin\sqrt{3/4L_{sn}C_p}\,t \quad (22)$$

The peak current and power will then occur at the time

$$t=\frac{2\pi}{3\sqrt{3}}\sqrt{L_{sn}C_p}$$

and will have the values

$$I_{n\,\max}=V_{cp\,\max}\sqrt{\frac{C_p}{L_{sn}}}\,\epsilon^{-\pi/3\sqrt{3}} \quad (23)$$

$$P_{\max}=(V_{cp\,\max})^2\frac{C_p}{L_{sn}}\,\epsilon^{-2\pi/3\sqrt{3}}R_L \quad (24)$$

The peak power resulting from equation 24 is approximately 10% larger than the peak power for critical damping as given by equation 20. This holds true if R_n is negligibly small in comparison to R_L.

The latter case of discharge was discussed because it represents the condition of a single section pulse forming network. In practice, pulse forming networks of five L-C (inductance-capacitance) sections are used. In this case, the load resistance is also made equal to the network impedance.

Design Considerations

The following main data are, in general, available:

$P_{\max}$ = peak power of output pulse, watts
t_p = pulse width, μsec
f_r = maximum pulse repetition frequency, cps (cycles per second)
$V_{\max}$ = peak voltage of output pulse, volts

An over-all efficiency η_m of the modulator must be assumed and a suitable value of the supply voltage V_{DC} must be determined. The value of η_m is based on previous experience. The voltage V_{DC} must be chosen so that neither the forward breakover voltage nor the peak inverse voltage of the controlled rectifiers of the first and second stage are exceeded.

The value of capacitor C_1 of the charging stage is found with fair approximation to be

$$C_1=\frac{P_{\max}t_p}{2V_{DC}^2\eta_m}\ \text{(microfarads)} \quad (25)$$

Inductance L_1 of the first stage may be determined as follows: The maximum time allowed for charging of capacitor C_1 is determined by the required maximum pulse repetition frequency f_r. This time should be somewhat (10%) smaller than the time $t_r=1/f_r$ elapsing between pulses, allowing complete discharge of the succeeding stages before a new charging of the first is initiated. This yields

$$t_1=\pi\sqrt{L_1C_1}=0.9\frac{1}{f_r}$$

or

$$L_1\cong\frac{0.81(10^3)}{\pi^2 f_r^2 C_1}\cong\frac{82}{f_r^2C_1}\ \text{(millihenrys)} \quad (26)$$

where f_r is given in kilocycles/sec and C_1 in microfarads (μf). The relatively long charging time t_1 keeps the peak power drawn from the supplying d-c source small. The charging time t_2 of stage II can be chosen as short as the surge current rating of the controlled rectifier CSR_2 will permit. Neglecting losses occurring in stage II the peak current resulting from the equation 11 will amount to

$$I_{2\,\max}=V_{c1\,\max}\sqrt{C_0/L_2}$$

where

$$C_0=\frac{C_1C_2}{C_1+C_2}$$

L_2 can be found from this relationship, time t_2 is then $t_2=\pi\sqrt{C_0L_2}$. The value of C_2 should be chosen so that equation 15 will be satisfied. Time compression ratios t_1/t_2 from 50 to 100 will be achievable in practice. This will result in a reduced number of stages and better efficiency when compared with the Melville modulator.

With regard to the large pulse currents

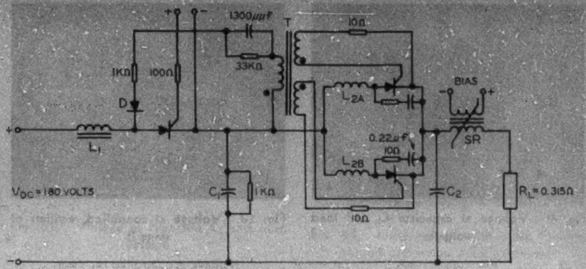


Fig. 5. Circuit of experimental modulator

L_1 = 2.5 millihenrys, Silectron C-core no. AJH114, 167 turns no. 20, 64 mil-AIRGAP
$L_{2A} = L_{2B}$ = 6.8 μhenrys, 13 turns no. 12 in 2 layers, D≅4 cm, L≅2.3 cm

SR: Deltamax core no. 6100-D1:
  Load winding: 6 parallel Litz wires of 7 strands no. 25, 4 turns
  Bias winding: 10 turns no. 24

C_1 = 4.25μf } 400-volt d-c paper capacitors
C_2 = 5.0 μf }

All Controlled Rectifiers: ZJ39A-300
T: Ferrite ring core, general ceramics, Type H, F-846-3
  ID = 0.281 inch, OD = 0.50 inch, H = 0.25 inch
  Primary 100 turns, secondary 2×50 turns
D: Silicon diode 1N458

occurring within the second stage it may be mentioned at this point that with different breadboard modulators peak charging currents of 80 amperes for a 15-μsec half-cycle and about 120 amperes for an 8-μsec half-cycle flowing through a single controlled rectifier, type *ZJ-39A* were measured. Because of limited facilities, repetition rates not higher than 500 cps were used during this investigation. No failures of controlled rectifiers because of these currents were experienced during breadboard evaluation.

Design principles for succeeding stages which employ saturable reactors as switching elements are given in sufficient detail in the existing literature. For a practical design it will be advisable to step up the voltage of the output capacitor C_2 of the second stage by means of an intermediate pulse transformer. The output transformer *PT* will then be designed further to step up the voltage to the values required for a magnetron or a traveling wave tube.

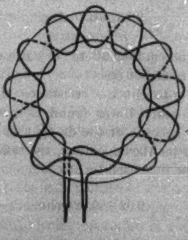


Fig. 6. Winding arrangement of saturable reactor

Further design considerations will be made in connection with the following discussion of an actual breadboard modulator, constructed and tested during the course of this work. The circuit of this modulator is pictured in Fig. 5. The modulator consists of three stages and employs two controlled rectifiers in parallel within stage II. Design values of components are given in Fig. 5. The R-C (resistance-capacitance) networks connected in parallel to the controlled rectifiers CSR_{2A} and CSR_{2B} were found necessary to suppress voltage spikes when these rectifiers interrupt flow of current. Since a finite time elapses until a controlled rectifier blocks flow of current, this current may overshoot zero to about 15% of its peak value and then be abruptly cut off. Thus, a large rate of change of current will exist and since this current flows through inductors L_{2A} and L_{2B} large reverse voltage spikes will appear across the controlled rectifiers. These spikes can be suppressed effectively by proper choice of the values of resistance and capacity. Best conditions are achieved if the shunting resistor is made about equal to the value of critical damping determined by either inductance and the shunting capacitor. Various R-C networks were tried for achieving an acceptable value of overshoot (approximately 10–15%) for the smallest capacitor required.

Special winding techniques were employed for the saturable reactor of the final stage. It is important that the ratio of unsaturated to saturated impedance be made as large as possible and that additional copper losses because of skin effect should be reduced to a minimum. To reduce the saturated inductance a winding arrangement as shown in Fig. 6 was used. With a conventional winding not only the number of turns wound around the core case contribute to the saturated inductance but there is "one turn" equal in diameter to the mean diameter of the core. Since the enclosed area of this turn is large and the path length for the flux relatively small, it could contribute considerably to the over-all saturated inductance of the reactor. This is especially true because the actual number of turns wound around the core must be kept small. With the winding arrangement of Fig. 6 the effective inductance of this one turn is cancelled. In the breadboard modulator only four turns wound on a Deltamax core no. *6100-D1* were necessary. To more evenly distribute the winding around the core and to reduce copper losses six windings in parallel were used.

Reduction of additional copper losses due to skin effect was achieved by using 7-strand no. 25 litz wires instead of solid wires. Total losses occurring in the final discharge stage were reduced by approximately 15%.

The type of capacitors must be care-

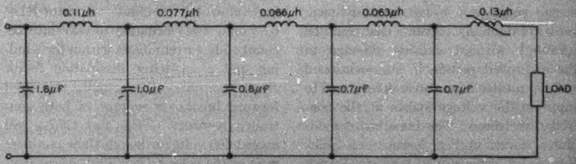


Fig. 7. Five-section pulse-forming network

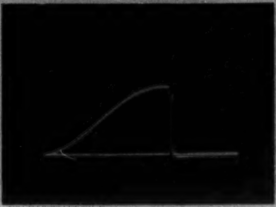


Fig. 8. Voltages at capacitors C_1 and C_2

V_{C_1} and V_{C_2}
100 volts/division
50 μsec/division

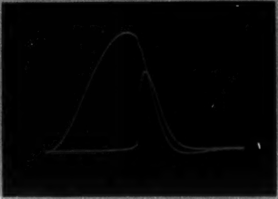


Fig. 9. Voltage at capacitor C_2 and load voltage

V_{C_2}—upper trace
V_L—lower trace
50 volts/division
2 μsec/division

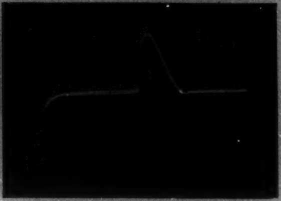


Fig. 10. Voltage at controlled rectifiers of stage II

Voltage at controlled rectifiers
CSR_{2A} and CSR_{2B}
100 volts/division
2 μsec/division

fully chosen to minimize losses within the modulator. Pyranol capacitors exhibit extremely high dissipation under the condition of pulse operation. It was found that standard 400-volt d-c paper capacitors operated much more satisfactorily.

The circuit of the breadboard modulator of Fig. 5 was later modified by replacing the output capacitor C_2 with a 5-section pulse forming network as illustrated in Fig. 7. Details for the design of such networks are given in reference 3. No particular effort for an ultimate design of this network was made since it was not the objective of this investigation and experiments were performed only on the basis of a simulated load. The saturated inductance of the saturable reactor within the final stage, which represents a portion of this network, should have been about 40% smaller than it actually happened to be. An output pulse transformer, to match the modulator to an available 51.5-ohm radio-frequency load, was also constructed and used during tests. A Deltamax core no. *4180-D1* was employed with a 4-turn primary and a 41-turn secondary winding. The primary winding was wound with four 7-strand no. 25 Litz wires in parallel. The transformer was furnished with a bias winding of 10 turns to reset its flux during interpulse periods.

Experimental Results

At the beginning of this investigation a few breadboard modulators using only one controlled silicon rectifier within the second stage were tested to study peak values of discharge currents that could be tolerated without causing damage to the controlled rectifiers. As mentioned earlier, suitable means were determined to suppress the voltage spikes at the controlled rectifiers. The breadboard modulator shown in Fig. 5 was then built. Results of experiments performed with this modulator will be reported here.

Since it had first been intended to operate the modulator up to pulse-repetition frequencies of 2,500 cps the first stage was designed for a relatively short charging time (about 0.33 millisecond). Based on preliminary experiments, it was estimated that an output pulse of approximately 0.15 watt-second energy could probably be achieved with a modulator using two controlled rectifiers within the second stage. If an over-all efficiency of $\eta=0.65$ and an input voltage $V_{DC}=165$ volts is assumed (with due regard to losses in the first stage), a voltage at capacitor C_1 not higher than 300 volts would result. Then the capacity C_1 resulting from equation 25 becomes

$$C_1=\frac{0.15(10^6)}{2(165^2)(0.65)}\cong 4.25\ \mu f$$

With a chosen $L_1=2.5$ millihenrys the charging time of stage I is

$$t_1=\pi\sqrt{2.5(4.25)(10^{-9})}=324\ \mu\text{sec}$$

A Silectron *C*-core no. *AJH114* was used for the inductor L_1 of the first stage, wound with 167 turns no. 20 wire. Unfortunately, only *C* cores of 4-mil lamination thickness were available. This resulted in relatively large core losses for this inductor at an apparent frequency of about 1,500 cps.

Resistor R_{fe}, see Fig. 4(A), representing these core losses was estimated from core loss characteristics to be approximately 500 ohms. The earlier mentioned shunting resistor R_p for the first capacitor C_1 had value of 1,000 ohms. Resistor $R1=0.84$ ohm was calculated to take into account eddy current losses within the winding and a capacitor dissipation factor $D=0.007$. Since the actually measured forward breakover voltage of both controlled rectifiers CSR_{2A} and CSR_{2B} will exceed 400 volts (although these rectifiers were of the type *ZJ39A-300*), it was considered safe to use an input voltage $V_{DC}=180$ volts instead of 165 volts during experimentation.

With these values, subtracting one voltage for rectifier drop from V_{DC} one obtains from equations 4, 3, and 7:

$$\alpha=\frac{1}{2}\left[\frac{840}{2.5}+\frac{1{,}500}{1{,}000(500)(4.25)(10^{-6})}\right]\cong 522\ \text{sec}^{-1}$$

$$I_{1\,\max}=179\left[\frac{1}{500}+\sqrt{\frac{4.25(10^{-3})}{2.5}}=\epsilon^{-522(162)(10^{-6})}\right]$$

$$=179[0.002+0.0412(0.919)]=7.15\ \text{amperes}$$

$$V_{C_1\,\max}=179(1+\epsilon^{-522(324)(10^{-6})})=330\ \text{volts}$$

Fig. 8 indicates the voltage at C_1 during the charging interval t_1, the voltage at C_2 during time t_2 is also shown as a narrow spike.

Inductors L_{2A} and L_{2B} of stage II were each designed for an inductance of 6.8 microhenrys. They were made of two layers of no. 12 stranded insulated wire with a total of 13 turns. The coils had a mean diameter of 4 cm (centimeters) and a length of about 2.3 cm. Capacitor C_2 had a value of 5 μf. The transmission time of stage II became

$$t_2=\pi\sqrt{\frac{4.25(5)}{9.25}(3.4)}=8.75\ \mu\text{sec}$$

which is equivalent to a natural frequency of 56,900 cps.

The apparent a-c resistance of both coils in parallel was found to be 0.056 ohm.[4] Because of the dissipation factor $D\cong 0.02$ for the capacitors, the resistance has to be increased by

$$\frac{10^6}{56{,}900(2\pi)2.3}(0.02)\cong 0.024\ \text{ohm}\ (C_0=2.3\mu f)$$

The series equivalent of the magnetizing impedance of the saturable reactor of

Vol.
Pt. 1
195

Fig. 11. Voltage at capacitor C_2 and load voltage

V_{C_2}—upper trace
V_L—lower trace
50 volts/division
1 μsec/division

Fig. 12. Load voltage with pulse forming network and output pulse transformer

Load voltage at $R_L=51.5$ ohms (RF-type) with pulse forming network and pulse transformer
400 volts/division
1 μsec/division

Fig. 13. Load voltage with pulse forming network

Load voltage at $R_L=0.315$ ohm with pulse forming network
50 volts/division
1 μsec/division

stage III has to be added to the sum of the foregoing values (about 0.013 ohm for an approximate magnetizing current of 6 amperes). Thus one obtains for R_2 [see Fig. 4(B)]:

$$R_2=0.056+0.024+0.013=0.093 \text{ ohm}$$

With the foregoing values the peak discharge current I_2 and the voltages V_{c_2} and V_{c_1} at the end of the transition interval t_2 can be found from equations 11, 13, and 14 as follows

$$\frac{R_2}{2L_2}=\alpha=\frac{0.093(10^6)}{2(3.4)}=13{,}650 \text{ sec}^{-1}$$

$$I_{2\,\max}\cong 329\sqrt{2.3/3.4}\,\epsilon^{-13650[8.75/2(10^{-6})]}$$

$$\cong 329(0.822)(0.942)\cong 255 \text{ amperes}$$

$$V_{c_2} \text{ (at time } t_2)\cong 329\left(\frac{4.25}{9.25}\right)\times$$

$$[1+\epsilon^{-13650(8.75)(10^{-6})}]$$

$$=151[1+0.888]\cong 285 \text{ volts}$$

$$V_{c_1} \text{ (at time } t_2)\cong 329\left(\frac{5}{9.25}\right)\times$$

$$\left[\frac{4.25}{5}-0.888\right]\cong -6.6 \text{ volts}$$

This minor value of voltage at C_1 at the end of time t_2 is of negligible importance, expecially if one considers uncertainties in determining the value of R_2.

The oscillogram of Fig. 9 shows the voltage at capacitor C_2 during the transition period. Peak value and time are in good agreement with the foregoing calculation. In its lower trace this oscillogram shows the voltage at the load of the modulator for nearly critical damping. The oscillogram of Fig. 10 exhibiting a double trace is a picture of the voltages occurring at controlled rectifiers CSR_{2A} and CSR_{2B} during operation of the modulator. Both voltages are nearly the same. The lower portion shows the voltage at the instant of firing while the upper portion indicates the reverse voltages. Winding and core data for the saturable reactor of the output stage have been given already in the section on "Experimental Results." With these data the calculated saturated inductance, assuming a relative permeability of $\mu_r=1$, becomes

$$L_s\cong 0.032 \ \mu\text{henry}$$

Actually it was found from measurements of time $t_p=\sqrt{L_sC_2}$ for the critically damped condition, that the actual saturated inductance amounted to

$$L_s\cong 0.13 \ \mu\text{henry}$$

The peak discharge current into the load was about 600 amperes. With four turns and a mean path length of 24 cm of core no. *6100-D1* the peak magnetizing force amounted to about 100 ampere-turns/cm. The measured value of L_s indicates that the average permeability was even with this relatively high magnetizing force still about $\mu_r\cong 4$. For the values of C_2 and L_s, the value of total resistance R_L+R_3 [see Fig. 4(C)] becomes for critical damping

$$R_L+R_3=2\sqrt{\frac{0.13}{5}}\cong 0.32 \text{ ohm}$$

For a rise time $t_p=\sqrt{0.13(5)}\cong 0.8$ μsec equivalent to a frequency $f\cong 313{,}000$ cps the a-c resistance of the saturable reactor winding was calculated as

$$R_{SR}\cong 0.003 \text{ ohm}$$

With an approximate dissipation factor $D=0.032$ for capacitor C_2, an equivalent series resistor R_d of

$$R_d=0.032\left[\frac{10^6}{313{,}000(2\pi)5}\right]=0.0032$$

has to be added to R_{SR} resulting in a total

$$R_3\cong 0.006 \text{ ohm}$$

A load resistor $r_L=0.315$ ohm was used. Using the foregoing values one obtains from equations 19 and 20

$$I_{3\,\max}=285\sqrt{5/0.13}(1/\epsilon)=650 \text{ amperes}$$

and

$$P_{\max}=650^2(0.315)=133 \text{ kw}$$

The peak value of the load voltage is therefore

$$V_{L_{\max}}=650(0.315)=205 \text{ volts}$$

The oscillogram of Fig. 11 shows, on an extended time scale of 1.0 μsec/division, the load voltage and the voltage at capacitor C_2 during the final discharge period. The load voltage can also be seen in the oscillogram of Fig. 9. From the maximum slope of the capacitor voltage, indicated by a straight line in Fig. 11, the peak discharge current can be determined as

$$I_{3\,\max}=\frac{\Delta v}{\Delta t}\,C_2=\frac{300}{2.5}(5)\cong 600 \text{ amperes}$$

resulting in a peak output power of

$$P_{\max}=600^2(0.315)\cong 113 \text{ kw}$$

Peak current and power determined from the voltage across the hairpin-shaped load resistor of 0.315 ohm are the same. Agreement with previously calculated values can be considered fair for this type of measurement if one also considers that the core losses of the saturable reactor during the interval of saturation had not been taken into account.

Fig. 12 shows an oscillogram of the load voltage, when capacitor C_2 had been replaced by the pulse forming network of Fig. 7. Averaging out the first small excursion at the start of the pulse, the peak power amounts to approximately

$130^2/0.315 \cong 53.5$ kw

The total pulse energy in watt-seconds transferred to the load is found to be about

$W_L = 0.17$ watt-second

From equation 25 the over-all efficiency of the modulator yields

$$\eta_m = \frac{W_L}{2(V_{DC})^2C_1} = \frac{0.17(10^6)}{2(180^2)(4.25)} \cong 62\%$$

Fig. 13 shows an oscillogram of the voltage across a standard radio-frequency load of 51.5 ohms, when this load was matched to the modulator by the output pulse transformer, described earlier, with a known ratio of 51/4. The peak power is approximately 50 kw. With a total energy of about 0.16 watt-second transferred to the load, the over-all efficiency of the modulator becomes

$\eta_m \cong 58\%$

If a more sophisticated design for a practical application of the modulator is used, even higher efficiencies can be achieved. Further reduction in output pulse duration will result in an increased output power. The total weight of components for the breadboard modulator including the output pulse transformer was 4 pounds.

During the course of experiments, one major modification of the basic circuit of Fig. 1 was investigated. The first controlled rectifier CSR_1 was replaced by an ordinary silicon rectifier and the parallel controlled rectifiers were fired directly from a trigger source. In general, the circuit performed as satisfactorily as expected. However, on several occasions this kind of circuit was not observed to be as fail-safe as the original circuit. Blowing of fuses, located in the incoming supply line, was experienced several times. This never occurred with the original circuit. The modulator operates most safely therefore if a controlled rectifier is used in the first stage since this arrangement effectively isolates the second stage from the line. Controlled rectifiers CSR_{2A} and CSR_{2B} cannot conduct unless controlled rectifier CSR_1 is turned off.

Conclusions

Experiments have demonstrated that it is possible to construct with presently available controlled silicon rectifiers small power radar modulators, capable of furnishing a pulse output power in the order of about 100 kw to the load. A pulse width of approximately 1.5 μsec is hereby assumed. Further development of controlled rectifiers to larger current ratings will probably result in peak output power ratings one order of magnitude larger than the foregoing value.

The reliability of such a modulator will be much improved in comparison to modulators employing thyratron tubes, eliminating the need for heating power.

A source of relatively low d-c voltage can be used to supply the modulator.

References

1. THE USE OF SATURABLE REACTORS AS DISCHARGE DEVICES FOR PULSE GENERATORS, W. S. Melville. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 98, pt. III, 1951, pp. 185–207.

2. MAGNETIC PULSE MODULATORS, K. J. Busch, A. D. Hasley, C. Neizert. *The Bell System Technical Journal*, New York, N. Y., Sept. 1955, pp. 943–93.

3. PULSE GENERATORS (book), A. H. Glascoe, J. V. Lebacqz. Radiation Laboratory Series, McGraw-Hill Book Company, Inc., New York, N. Y., vol 5, 1948.

4. RADIO ENGINEERS HANDBOOK (book), F. E. Terman. McGraw-Hill Book Company, Inc., 1943, pp. 77–81.

Vol.
Pt. 1
1959

Analysis of Magnetic Amplifiers Using a Resistive Reactor Model

SHIMSHON FRANKENTHAL
NONMEMBER AIEE

THE LACK of a simple and reasonably accurate model for the behavior of magnetic materials has been the major source of difficulty in all the attempts to analyse and predict the behavior of magnetic amplifiers. In several cases, even a qualitative understanding of the operation of certain circuits was impossible to achieve because of the lack of such a model.

However, upon reviewing some of the past efforts in this area, one is struck by the fact that some of the neatest analyses of the performance of relatively simple amplifiers were made possible by the use of a resistive model to represent the behavior of the reactor.[1,2] One is, therefore, tempted to explore the following questions: First, can such a model lead to a quantitative understanding of the operation of relatively complicated magnetic amplifiers, such as the polyphase circuits? Second, can it be used to predict the performance characteristics of these amplifiers in terms of the properties of the reactor core material? The latter question cannot be answered intelligently unless one is familiar with the assumptions and approximations involved in the development of the resistive model from ferromagnetic theory, and this report will therefore open with an outline of one possible development of this sort.

Development of the Resistive Model

Woodson has described the development of a simplified mathematical model for toroidal thin-tape reactor cores.[3] One of the conclusions of that work is that the rate of change of flux in a reactor which is in the midst of a flux-reversal process may be expressed as

$$\frac{d\varphi}{dt} = 2B_s g(r) \frac{dr}{dt} \tag{1}$$

In this expression, B_s represents the saturation flux-density; whereas, r represents an averaged and normalized domain-wall dimension, which essentially measures the position of an "average" domain wall. The function $g(r)$ indicates the rate of growth of the reverse-saturated region with respect to the position of this average domain wall. Woodson finds that, in toroidal Orthonol cores with a high ID (inside diameter) to OD (outside diameter) ratio, the average domain-wall velocity dr/dt is approximately related to the intensity H of the

---

Paper 59-1046, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-IRE Nonlinear Magnetics and Magnetic Amplifiers Conference, Washington, D. C., September 23–25, 1959. Manuscript submitted November 12, 1958; made available for printing July 16, 1959.

SHIMSHON FRANKENTHAL is with the Massachusetts Institute of Technology, Cambridge, Mass.

This paper is taken in part from a thesis submitted by the author to the Massachusetts Institute of Technology in partial fulfillment of the requirement for a Master of Science degree. The work was supported in part by the Wright Air Development Center under Air Force contract *AF33(616)-3242*.

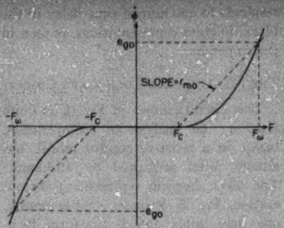


Fig. 1. Rate of change of flux versus applied mmf

applied magnetic field by the following expression

$$\frac{dr}{dt}=K(H-H_c)^2 \tag{2}$$

where K is a material constant and H_c the d-c coercive force. Knowing that such a reactor will maintain a constant field intensity H when normally excited by a sinusoidal voltage, one may deduce the form of the function $g(r)$ to be

$$g(r)=\frac{\pi}{2}A_c\sin\pi r \qquad 1>r>0 \tag{3}$$

where A_c is the effective cross-sectional area of the core. As a by-product of this development, the relation between the a-c loop width H_ω and the frequency of excitation ω is found to be

$$H_\omega-H_c=\sqrt{\frac{1}{K\pi}}\,\omega \tag{4}$$

A similar relation exists between the average width of the hysteresis loop along which the core operates when excited by a square-wave voltage.

Some simplification of these equations may be achieved if one approximates the sinusoidal function $g(r)$ by its average $g_a(r)$ over the range $0<r<1$. One thus obtains a rectangular pulse whose amplitude is

$$g_a(r)=A_c \qquad 1>r>0 \tag{5}$$

Equation 1 then simplifies to read

$$\frac{d\varphi}{dt}=2KB_sA_c(H-H_c)^2 \tag{6}$$

implying that the reactor will require a constant field intensity to maintain a *constant* rate of change of flux. This is equivalent to the statement that the rate of change of flux at any moment depends only on the value of H at that moment. In their "exact" form, equations 1 through 3 imply that there is also a dependence on the past history of H, which is embodied in $g(r)$.

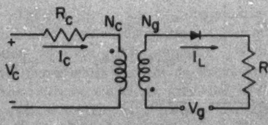


Fig. 2. Half-wave magnetic amplifier

Equation 6 is plotted in Fig. 1, with magnetomotive force (mmf) F replacing field intensity H. It is known that a reactor incorporated, say, in a full-wave self-saturating circuit may operate at different points on this curve during its reset period, even if the amplifier is controlled by a direct voltage and excited by a square-wave gate-supply voltage. In order to allow the calculations to be based on average control currents and rates of change of flux, one would prefer to approximate the parabolic relation of Fig. 1 by a straight line. One possible choice of such an approximation is indicated in dashed lines in Fig. 1. It is arrived at by recognizing that a reactor may be operated anywhere between the point of no reset ($\dot{\varphi}=0$, $f=\pm F_c$) and the point of full reset ($\dot{\varphi}=\pm\dot{\varphi}_{\max}$, $f=\pm F_\omega$). When operated at the latter point, the reactor is carried from positive to negative saturation during the reset interval, and the magnitude of its average rate of reset $\dot{\varphi}_{\max}$ is given by the number of volts-per-turn required for the so-called "normal" excitation. If the duration of the reset interval is one-half cycle of the gate-supply frequency ω, then

$$\dot{\varphi}_{\max}=e_{go} \tag{7}$$

where

$$e_{go}\triangleq\frac{2}{\pi}\omega B_sA_c \tag{8}$$

In this case, the slope of the linear approximation indicated in Fig. 1, which, dimensionally, is a resistance, is given by

$$r_{mo}=\frac{e_{go}}{F_\omega-F_c}=\frac{2}{\pi}-\frac{B_s}{H_\omega-H_c}\cdot\frac{A_c}{l_c}=2\sqrt{\frac{K}{\pi}}\,B_s\frac{A_c}{l_c}\sqrt{\omega} \tag{9}$$

where l_c is the mean length of the magnetic path. Thus, a piecewise linear approximation of the curve of Fig. 1 leads to a resistive model for the reactor, which consists of the magnetizing resistance shunted by a current source and several diodes to take care of the break-points.

Needless to say, other linear approximations are possible. It is, however, impossible to determine *a priori* which one, if any, is the best. One may therefore just as well accept the one presented here for the purpose of the analysis.

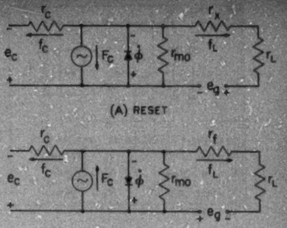


Fig. 3. Equivalent circuits for half-wave magnetic amplifier

Before proceeding to describe the results of the analyses which employ this model, it may be well to indicate briefly the way in which it enters the analysis and study some of its implications. Consider, then, the half-wave circuit, shown in Fig. 2, which is controlled by a direct voltage V_c and supplied with a square voltage of amplitude V_g. To simplify the construction of equivalent circuits for this amplifier, it is convenient to normalize the voltage, current, and resistance present in each winding with respect to the number of turns N in that winding, by letting

$$e=\frac{V}{N},\ f=NI,\ r=\frac{R}{N^2} \tag{10}$$

Remembering that, in practical cases, the control voltage is negative (it should tend to reset the reactor), one first constructs the model representing the left branch ($f<0$) of the flux-rate versus mmf curve and inserts it instead of the reactor. Theoretically, one should introduce an ideal 1:1 transformer to decouple the control and gate circuits; but the elimination of this transformer is permissible in most cases. One thus obtains an equivalent circuit for the reset interval. The equivalent circuit for the prefiring portion of the gating interval, during which the reactor is being set, may be obtained by inverting the direction of the current source and the diode incorporated in the model, in order to simulate the right-hand branch of Fig. 1. In this case, one lumps the series resistance of the gate winding and the resistance of the conducting gate rectifier in the symbol r_f; see Fig. 3.

During the reset interval, the reactor presents its magnetizing resistance to the control circuit, provided that the reverse resistance of the gate rectifier is infinite. The rate of change of flux is then deter-

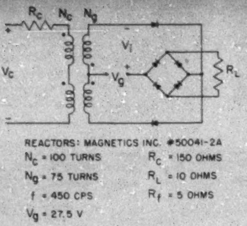


Fig. 4. Single-phase full-wave amplifier

mined solely by the voltages present in the control circuit, which are simply shared between r_c and r_m. Note that, after the diode incorporated in the model blocks, as it should if reset is to take place, the effective control voltage is

$$e_c' = e_c + r_c F_o \tag{11}$$

During the prefiring portion of the gating interval, the control voltage has little effect on the rate of change of flux, provided that the condition

$$r_f + r_g << \frac{r_m r_c}{r_m + r_c} \tag{12}$$

is satisfied. The rate of change of flux is then determined by the gate-supply voltage, which is almost totally impressed across the reactor. As far as the control circuit is concerned, the setting reactor reflects a voltage source having virtually no series resistance.

After saturation, the reactor is properly represented by a short circuit, which decouples the windings. The gate-supply voltage is then shared by the load resistance r_L and the total series resistance r_f. It is now convenient to define the internal voltage e_i, which is the amount by which the gate-supply voltage e_g exceeds the reactor voltage φ_{sat} during the gating interval. Averaging over the gating interval, one obtains the relation

$$\overline{e_i} = e_g - \overline{\varphi_{sat}} \tag{13}$$

in which the bar denotes the averaging operation.

This information is sufficient to allow the analysis of magnetic amplifiers with only rare need actually to draw an equivalent circuit. The actual procedure of the analysis is discussed in the Appendix, and its application to the amplifiers discussed in this work is relatively straightforward.[2,4] Suffice it to state, at this point, that the analysis leads to block-diagram representations which relate the average control voltage to the average internal and output voltage. The averaging is performed over suitably chosen intervals, the selection of which will be discussed later. In their exact form, these representations involve discrete time delays and require the use of difference equations, or Z transforms. The transition from this kind of representation to continuous-data "language" is simple when the signal frequencies are low compared to the gate-supply (carrier) frequency.[5]

Analysis and Experimental Observations

In this section, the results of a theoretical study of the operation of three self-saturating magnetic amplifiers are presented and compared with experimental evidence. The normalized variables defined in equation 10 are used throughout this discussion, being both illuminating and economical in space. The symbols r_m and F_o are adopted to designate the magnetizing resistance and the current source which constitute the reactor model. These will be related to r_{mo} and F_e as the need arises.

The assumptions and definitions involved in the analysis are the following:

1. The waveform of the gate-supply voltage is square, and its amplitude is the one required for normal excitation of the reactors.
2. The reverse resistance of the gate rectifiers is negligibly large compared to the magnetizing resistance. However, the total series resistance r_f of the gate circuit may be comparable with the load resistance r_L.
3. The control voltage remains constant throughout any representative interval. The duration of this interval is T/k for a circuit which incorporates k reactors and excited by a gate-supply voltage whose period is T. The same duration also corresponds to one unit of time delay in the block-diagram representations as seen in the Appendix.

THE SINGLE-PHASE FULL-WAVE CIRCUIT

This circuit, shown in Fig. 4, is quite familiar; an analysis of its operation, based on a resistive reactor model, has already been presented by Hubbard.[2] The block diagram representing its operation, Fig. 5, is linear; consequently, one may apply the Z-transform technique to obtain an exact transfer function

$$\frac{\Delta e_i(Z)}{\Delta e_c(Z)} = \frac{r_m}{r_c} \cdot \frac{Z}{1 + \frac{r_m}{r_c}(1 - Z)} \tag{14}$$

relating increments of the internal voltage and the control voltage within the range of control of the amplifier. At sufficiently low signal frequencies, one may approximate

$$Z = e^{-\frac{T}{2}s} \sim \frac{1}{1 + \frac{T}{2}s} \tag{15}$$

in which case the transfer function becomes

$$\frac{\Delta e_i(s)}{\Delta e_c(s)} = G_i \frac{1}{1 + \tau s} \tag{16}$$

where the internal voltage gain G_i is

$$G_i = \frac{r_m}{r_e} \tag{17}$$

and the time constant τ is

$$\tau = \left(1 + \frac{r_m}{r_c}\right)\frac{T}{2} \text{ sec} = \left(1 + \frac{r_m}{r_c}\right)\frac{1}{2} \text{ cycles} \tag{18}$$

It is interesting to note that an exact step response of this amplifier, which may be calculated from equation 14, will involve transients of the form

$$\left(\frac{r_m}{r_c + r_m}\right)^n$$

where n is the sequence number of the half cycle. If one allows $t = nT/2$, this form may be re-expressed as an exponential

$$\left(\frac{r_m}{r_c + r_m}\right)^n = e^{\frac{2}{T} \ln \frac{r_m}{r_c + r_m} t}$$

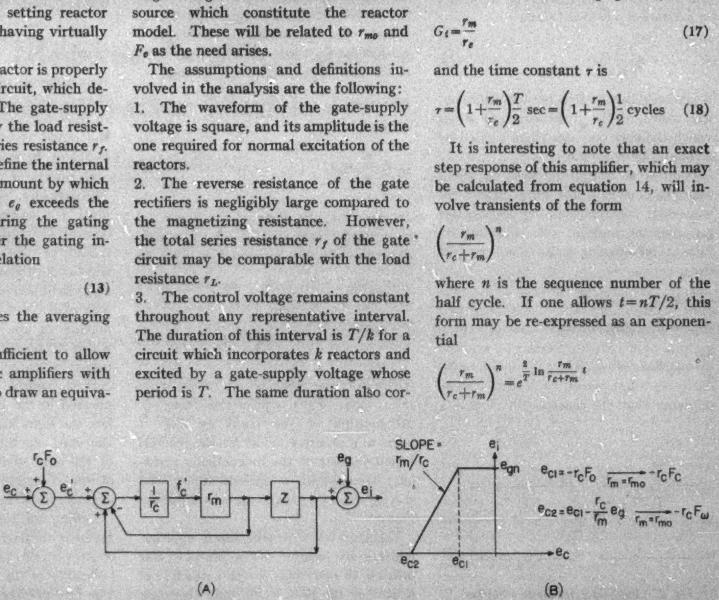

r_cF_o
e_c
e_c'
Σ
1/r_c
f_c'
r_m
Z
e_g
e_i
(A)
SLOPE = r_m/r_c
e_i
e_gn
e_c2
e_c1
e_c
e_c1 = -r_cF_o → (r_m = r_mo) -r_cF_c
e_c2 = e_c1 - (r_c/r_m) e_g → (r_m = r_mo) -r_cF_ω
(B)

Fig. 5. Single-phase full-wave amplifier

A—Block-diagram representation

B—Control characteristic

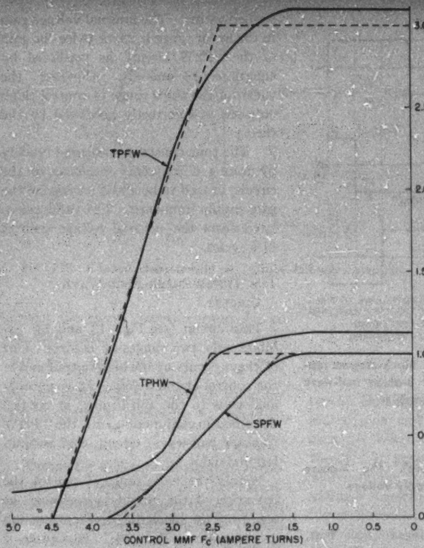


Fig. 6 (left). Experimental control characteristics of various amplifiers

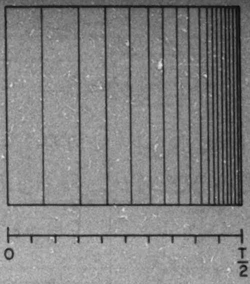


Fig. 7 (right). Step response of single-phase full-wave amplifier

THE VERTICAL LINES INDICATE THE SATURATION-TIME OF A REACTOR DURING SUCCESSIVE CYCLES OF THE TRANSIENT.

One may now identify a time constant τ' such as

$$\frac{1}{\tau'}=-\frac{2}{T}\ln\frac{r_m}{r_c+r_m}=\frac{2}{T}\ln\left(1+\frac{r_c}{r_m}\right)$$

and recognize that, when $r_m >> r_c$, both τ' and τ, as given by equation 18, approach the limit

$$\frac{r_m}{r_c}\frac{T}{2}$$

The reset period of this circuit is one-half cycle. Under minimum output conditions, the magnitude of the average rate of reset is equal to the magnitude of the gate-supply voltage. Thus, for normal excitation, this amplitude should be

$$e_{gn}=e_{go} \tag{19}$$

where e_{go} is given by equation 8. One may now identify the magnetizing resistance r_m with r_{mo} as given by equation 9, and recognize that linear control of the amplifier is possible for control voltages which lie within the range

$$-r_cF_\omega<e_c<-r_cF_c \tag{20}$$

The analysis is thus seen to provide a block diagram which relates the gain and time constant of the amplifier. If one further assumes that the development of the resistive model, as presented in the previous section, is legitimate, one may expect to be able to predict the value of the gain, as well as the shape of the control characteristic, from the available data about the reactor core material. The objective of the experimental work is to check both.

The parameters of the circuit used in the experimental work are given in Fig. 4. The experimental control characteristic, plotting the normalized internal voltage versus the control mmf, is shown in Fig. 6. The transient response to a step of control voltage, which drove the output from maximum to minimum, is shown in Fig. 7. The pertinent experimental observations may be summarized as follows:

1. The control characteristic is linear over most of the range of control. The internal voltage gain in the linear region is 11.5, and the magnetizing resistance calculated from it is 0.173 ohms per turn.[2]
2. The observed time-constant is 6 cycles; see Fig. 7. The time constant predicted by the analysis, $G_i/2$ cycles, is 5.75 cycles.
3. The limits on the range of control, in terms of the control mmf, are 1.35 and 3.8 ampere-turns. The limits on the extrapolated linear portion of the control characteristic, the dashed lines in Fig. 6, are 1.6 and 3.6 ampere-turns. The lower limit of either set agrees reasonably well with the d-c coercive mmf of 1.44 ampere-turns, calculated from manufacturer's data. Manufacturer's data about the a-c loop width at 60 cps (cycles per second) and 400 cps were used, in conjunction with equation 4, to calculate a value of 3.8 ampere-turns for the a-c loop width at the operating frequency (450 cps).

THE THREE-PHASE HALF-WAVE CIRCUIT

The analysis of this circuit (see Fig. 8) is complicated by the fact that it exhibits three ranges of control (see Figs. 6 and 11). Within the first range, corresponding to internal voltages

$$0<e_i<\frac{e_g}{2}$$

the operation of the amplifier is nonlinear; and its gain is rather low. The block diagram pertaining to this range is indicated in Fig. 9(A). This is a time-domain representation, and the symbol Z appearing in it should not be interpreted as a transform variable. It merely denotes a time-delay of one third of a cycle.

Within the second range of control, corresponding to internal voltages

$$\frac{e_g}{2}<e_i<e_g$$

the circuit operates in a linear fashion.

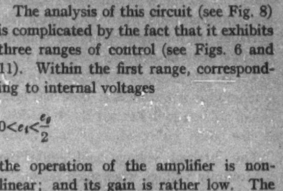
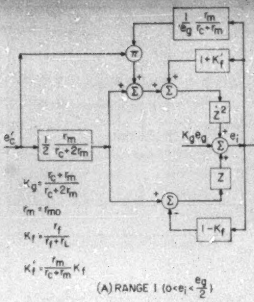
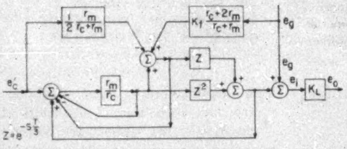


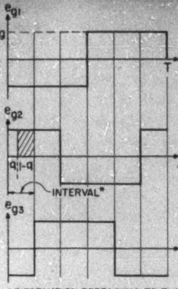
Fig. 8. The 3-phase half-wave amplifier


$\frac{1}{e_g}\frac{r_m}{r_c+r_m}$
π
$1+K_f'$
Z^2
e_c'
$\frac{1}{2}\frac{r_m}{r_c+2r_m}$
$K_g e_g$
e_i
K_L
e_o
Z
$1-K_f$
$K_g = \frac{r_c+r_m}{r_c+2r_m}$
$r_m = r_{mo}$
$K_f = \frac{r_f}{r_f+r_L}$
$K_f' = \frac{r_m}{r_c+r_m}K_f$

(A) RANGE I $(0<e_i<\frac{e_g}{2})$



(B) RANGE II $(\frac{e_g}{2}<e_i<e_g)$



*INFORMATION PERTAINING TO THE ANALYSIS OF THE TPFW AMPLIFIER:
q - PREFIRING
1-q - POSTFIRING

Fig. 9 (left). Block-diagram representation of 3-phase half-wave amplifier

Fig. 10 (above). The 3-phase gate-supply voltage

An exact transfer function may be obtained from the block diagram of Fig. 9(B). However, if one assumes $r_m>>r_c$ and considers sufficiently low signal frequencies, the following simplified transfer function may be derived:

$$\frac{\Delta e_i(s)}{\Delta e_c(s)} \sim 2\frac{r_m}{r_c}\cdot\frac{1}{1+\frac{r_m}{r_c}\cdot Ts}\cdot r_c \tag{21}$$

The internal voltage will equal the gate-supply voltage when each reactor is saturated for one third of a cycle; see the waveforms of Fig. 10. Under this condition, a continuous current

$$\frac{e_I}{r_f+r_L}$$

is delivered to the load. If the resistance r_f is not zero, the magnitude of the control voltage e_{c3} required to produce this condition is somewhat higher than r_cF_o; see Fig. 11. As the magnitude of the control voltage decreases below e_{c3}, the firing angle of each reactor progresses. During part of the time, the load current is delivered through one saturated reactor having a series resistance r_f; and during the other part, it is delivered through two simultaneously saturated reactors in parallel. As a result, the average output current varies somewhat with the control voltage, until the latter reaches the value of $-r_cF_o$, at which time each reactor is saturated half the time. The gain in this range is sufficiently low, and the variation of output current sufficiently small, that it may be ignored completely. It is, however, gratifying to note that, as r_f approaches zero, the value of e_{c3} approaches $-r_cF_o$, so that this range of control disappears completely, as it should. Note also that r_f has no effect on the internal voltage gain in the linear range of control; it just shifts it bodily to the left as shown in Figs. 11 and 6.

Under minimum output conditions, reset is achieved at a uniform rate e_g during one half cycle. The normal excitation voltage of this circuit is again e_{go}, as given by equation 8.

The circuit used in checking these results incorporated the same components used in the experimental SPFW (single-phase full-wave) circuit (shown in Fig. 4). The experimental observations may be summarized as follows:

1. The control characteristic in Fig. 6 exhibits the three ranges of control described earlier. The internal voltage gain in the linear range is 23, or twice the gain of the SPFW circuit, as predicted by equations 17 and 21. However, the width of the third range of control (high outputs) is incorrectly predicted by the theory.
2. The time constant, measured crudely by taking a frequency response of the circuit, proved to be about 9 cycles of the gate-supply frequency. The value calculated from the internal-voltage gain is 11.5 cycles.

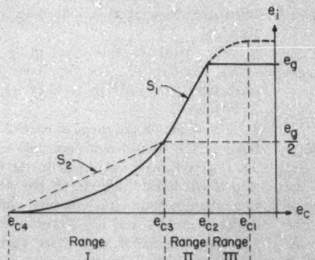
The Three-Phase Full-Wave Circuit

This circuit (see Figs. 12 and 13) exhibits only two ranges of control. One of them occurs at the high-output end of the control characteristic. It is entirely analogous to the third range of control described in conjunction with the TPHW (3-phase half-wave) circuit, and is similarly trivial.

Within the main range of control, the operation of the circuit is nonlinear, as seen from the block diagram representing its operation, Fig. 14. However, if $r_m>>r_c$, the deviation from linearity is very small; and one may readily derive the following approximate transfer function; see the Appendix

$$\frac{\Delta e_I(s)}{\Delta e_c(s)} \sim \frac{r_m}{r_c}\cdot\frac{1}{1+\frac{r_m}{r_c}\frac{T}{2}s} \tag{22}$$

This expression is identical in form with the transfer function of the SPFW circuit. However, the mode of operation of the circuit is such that, under minimum output conditions, each reactor is reset at a rate $2e_g$ during one sixth of a cycle as seen in Fig. 15. Thus, the normal excitation voltage for the circuit is 1.5 e_{go}; and the maximal rate of reset has a magnitude 3 e_{go}. If the parabolic relation between the flux rate and the mmf, as plotted in Fig. 1, is correct, one may



$$S_1 = \frac{r_m(3r_c+4r_m)}{2r_c(r_c+r_m)} \xrightarrow[r_m>>r_c]{} 2\frac{r_m}{r_c}$$

$$S_2 = \frac{1}{2}\frac{r_m(3r_c+4r_m)}{2(r_c+r_m)(r_c+2r_m)} \xrightarrow[r_m>>r_c]{} \frac{1}{2}$$

$$e_{c1} = -r_cF_o$$

$$e_{c2} = -r_cF_o - \frac{r_c(r_c+2r_m)}{r_m(3r_c+4r_m)}K_f e_g$$

$$e_{c3} = e_{c2} - \frac{e_g}{2S_1}$$

$$e_{c4} = -\left(1+\frac{r_c}{r_m}\right)e_g$$

Fig. 11. Control characteristic of 3-phase half-wave amplifier

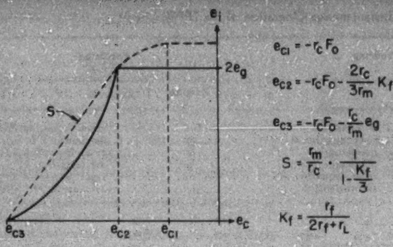


Fig. 12. Control characteristic of 3-phase full-wave amplifier

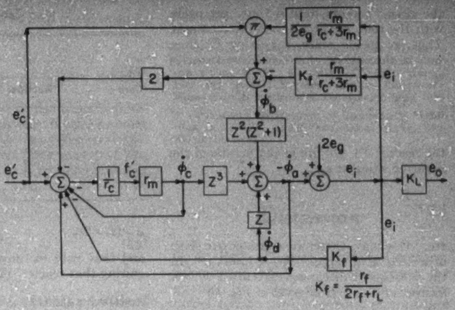


Fig. 14. Block-diagram representation of the 3-phase full-wave amplifier

expect the cutoff control current F_ω' to be such that

$$F_\omega' - F_c = \sqrt{3}(F_\omega - F_c) \qquad (23)$$

At the same time, the proper value for the magnetizing resistance r_m should be $\sqrt{3}r_{mo}$

Except for a change in the load resistance, the components of the experimental circuits were again the ones given in Fig. 4. The experimental control characteristic shown in Fig. 6 is the result of normalizing the internal voltage and taking account of the fact that its nominal maximum value is 3 e_{go}. The experimental observations may be summarized as follows:

1. The internal voltage gain over the linear portion of the range of control is 34. The value predicted on the basis of the gain of the SPFW circuit is 20. Correspondingly, the observed value of the cutoff control mmf, 4.5 ampere-turns, is remote from the value, 5.5 ampere-turns, predicted by equation 23.
2. The observed time constant is 15 cycles of the gate-supply frequency. Using the observed value of the internal voltage gain, equation 22 yields a "theoretical" value of 17 cycles.

Conclusions

A review of the experimental observations reported in the previous section reveals that the relation between the observed gain and time constant of each circuit is indeed the one predicted by the theory. It appears, then, that the resistive model is indeed capable of providing a correct quantitative understanding of the mechanism of operation of magnetic amplifiers, embodied here in the block-diagram representations for these amplifiers. Unfortunately, no such statement can yet be made regarding the prediction of the gain and time constant in terms of the core-material properties.

With regard to this question, it will be noted that the theoretically derived model does indeed predict the shape of the control characteristic, hence, the gain, of the SPFW circuit, as well as the gain of the TPHW circuit. This is, perhaps, not surprising. The reactors in both amplifiers operate over the same range of flux-reset rates, so that any linear approximation should hold equally well in both cases. The exact nature of the relation between the reset-rate and the mmf, which is basically dependent only on equation 2, is immaterial in this case, provided that it is amendable to a linear approximation, which apparently it is.

However, in expecting a similar prediction of the gain of the TPFW (3-phase full-wave) amplifier, whose reactors operate over a wider range of reset rates, one is actually testing the validity of the relation between the reset rate and the mmf or, specifically, of equation 2. The discrepancy between the predicted and the observed gains casts some suspicion on this relation. There already exists a certain amount of conflicting experimental evidence pertaining to it, and it therefore appears that any attempt to settle the discrepency at this time may be premature.[6,7] Nevertheless, one may expect that, once a more refined understanding of the relation between the reset rate and the mmf is achieved, a fully reliable utilization of the resistive reactor model, along the lines described here, will be possible.

Appendix

Analysis of the Three-Phase Full-Wave Amplifier

In studying the operation of this circuit in Fig. 13, it is advantageous to associate each reactor with its gate-supply waveform. The reactors $X1$, $X2$, and $X3$ are associated with the waveforms e_{g1}, e_{g2}, and e_{g3}, respectively (see Fig. 10). The $Y1$ reactors are associated with the negatives of these waveforms. Symmetry considerations then lead to the conclusion that the pattern of operation of

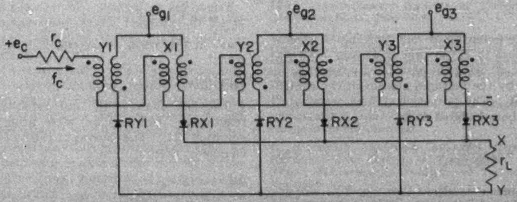


Fig. 13. The 3-phase full-wave amplifier

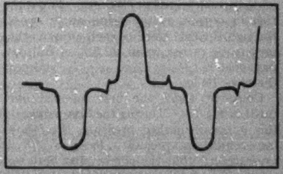


Fig. 15. Voltage across a reactor in 3-phase full-wave amplifier at minimum output

the circuit repeats itself 6 times per cycle, with the roles of the reactors interchanged. Thus, at steady state, if reactor $X1$ fires, say, in the middle of the positive swing of e_{g1}

$$t_{fX1}=\frac{3T}{4}$$

the reactor $Y1$ will fire in the middle of the negative swing of e_{g1}

$$t_{fY1}=\frac{T}{4}$$

etc. One may, therefore, consider one sixth of a cycle as a representative interval. In the subsequent discussion, attention will be focused on the first interval of Fig. 10

$$\frac{T}{6}>t>0$$

One may further state that, in order that an output current pulse may flow, at least one X reactor and one Y reactor must be saturated simultaneously. If only one reactor, say $Y3$, were saturated, any appreciable current through the load would be prevented by reactors in which the flux is changing.

Thus, if firing takes place at the beginning of the last third of each gate-voltage swing

$$t_{fX1}=\frac{5T}{6}\ t_{fY1}=\frac{T}{3},\ \text{etc.}$$

no output current can flow. As the firing angle sweeps through the middle third of each swing, the output increases. If the resistance r_f of the windings and rectifiers were zero, maximum output would be attained when the firing angle reached the beginning of the middle third of each swing

$$t_{fX1}=\frac{2T}{3},t_{fY1}=\frac{T}{6},\ \text{etc.}$$

since a voltage $2e_g$ would then be continuously switched on to the load. Since r_f is not zero, the actual output voltage at this point is

$$\frac{2r_L}{2r_f+r_L}e_g$$

and the maximum voltage

$$\frac{2r_L}{\frac{3}{2}r_f+r_L}$$

will be reached when the firing angle reaches the beginning of each swing, at which point two X reactors and one Y reactor (or vice versa) are saturated at all times. The variation in output as the firing angle sweeps through the first third of each swing is thus seen to be rather small, 12.5% of full output, even if r_f is as large as $r_L/2$, and may therefore be ignored.

Consider, then, the first interval indicated in Fig. 10. During the first portion of this interval, marked prefiring, the following conditions prevail: Reactor $Y3$ is saturated, so that the potential at both terminals X and Y is only slightly higher than $-e_g$. It is, therefore, clear that the gate rectifier $RX2$ will unblock, allowing the reactor to be set at a rate

$$\dot{\phi}_{X2}{}^1(t)=2e_g \tag{24}$$

and that gate rectifier $RY2$ will block requiring that reactor $Y2$ be reset at a rate

$$\dot{\phi}_{Y2}(t)=-r_m f_c'^1(t) \tag{25}$$

where

$$f_c'=f_c+F_o \tag{26}$$

The status of the gate rectifiers $RX1$, $RX3$, and $RY1$ is yet undetermined, since e_{g1}, e_{g3}, e_x, and e_y are almost identical. One may settle this question by the following reasoning: Suppose that rectifier $RX1$, for example, is blocked. The reactor $X1$ should then be reset by the current present in the control circuit. However, any substantial rate of reset in the reactor will establish a voltage across it which would tend to unblock the gate rectifier, thus violating the initial assumption. One is therefore forced to conclude that the rectifier actually conducts. A negligible resetting voltage e_x-e_{g1} is established across its gate winding, and insofar as the control circuit is concerned, it serves as a pure voltage source (short circuit). The same reasoning applies to all other rectifiers in question.

A bird's-eye view of the behavior of the various components is provided by Table I, in which information about the state of the gate rectifiers and the rate of flux change in the associated reactors is presented in a concise fashion.

Table I. Instantaneous Operation of the TPFW Circuit

	Prefiring (q)		Postfiring ($1-q$)	
Core	**Rectifier**	$\dot{\phi}$	**Rectifier**	$\dot{\phi}$
$X1$	on	no change	off	reset (equation 33)
$X2$	on	set (equation 24)	on	saturation
$X3$	on	no change	off	reset (equation 33)
$Y1$	on	no change	on	set (equation 31)
$Y2$	off	reset (equation 25)	off	reset (equation 33)
$Y3$	on	saturation	on	saturation

At some time t_f reactor $X2$ saturates, allowing a pulse of voltage whose magnitude $2e_g$ is now available to produce a load current. This available voltage is defined as the internal voltage e_i of the circuit. If the prefiring portion occupies a fraction q of the interval, then the average internal voltage during the interval is

$$\bar{e}=(1-q)2e_g \tag{27}$$

The actual load voltage is then

$$\bar{e}_o=K_L\bar{e}_i \tag{28}$$

where

$$K_L=\frac{r_L}{2r_f+r_L} \tag{29}$$

The potential at load terminal X is now up close to e_g, so that $RX1$ and $RX3$ unblock, allowing the associated reactors to be reset by the voltages present in the control circuit. The potential at load terminal Y may be calculated to be

$$e_y{}^2(t)=e_g\frac{r_L}{2r_f+r_L} \tag{30}$$

and since it is higher than e_{g1}, $RY1$ will unblock and the associated reactor will be set at a rate

$$\dot{\phi}_{Y1}{}^2(t)=e_y-e_{g1}=2K_fe_g \tag{31}$$

where

$$K_f=\frac{r_f}{2r_f+r_L} \tag{32}$$

The voltage reflected across this reactor into the control circuit will aid the control voltage in resetting reactors $X1$, $Y2$, $X3$. Since each of these reactors now presents its magnetizing resistance r_m to the control circuit, the total resistance in that circuit is r_c+3r_m. The rate of reset in each reactor is, then,

$$\dot{\phi}_{X1}{}^2(t)=\dot{\phi}_{X3}{}^2(t)=\dot{\phi}_{Y2}{}^2(t)=\frac{r_m}{r_c+3r_m}\times[e_c'(t)-2K_fe_g] \tag{33}$$

This completes the detailed description of the behavior of the circuit. Inspection of Table I now reveals that there are four significant modes of flux change. The first is associated with the gating reactor ($X2$ in the particular interval under consideration). The average rate of change of flux $\dot{\phi}_a(n)$ associated with this mode is such that

$$e_i(n)=2e_g-\dot{\phi}_a(n) \tag{34}$$

The second mode of flux change is the one exhibited by reactors $X1$ and $X3$, which are subjected to postfiring reset, at a rate given by equation 33. The average rate of reset in this mode is

$$\overline{\dot{\phi}_b}=\frac{r_m}{r_c+3r_m}[e_c'(t)-2K_fe_g](1-q) \tag{35}$$

and using equation 27 for q, one gets

$$\dot{\phi}_b(n)=\frac{r_m}{r_c+3r_m}\frac{e_i(n)}{2e_g}e_c'(n)-\frac{r_m}{r_c+3r_m}K_fe_i(n) \tag{36}$$

The third mode is the one assumed by reactor $Y2$, which is reset throughout the interval with its gate rectifier blocked. The instantaneous rate of reset given by the resistive model

$$\dot{\phi}(t)=r_m(f_c(t)-F_o)=r_mf_c' \tag{37}$$

may, in this case, be averaged directly to read

$$\dot{\phi}_c(n)=r_mf_c'(n) \tag{38}$$

The last mode is the one exhibited by reactor $Y1$, which is set after firing; see equation 31. The average rate of flux change is, then,

$$\dot{\phi}_d(n)=2K_fe_g[1-q(n)]=K_fe_i(n) \tag{39}$$

In addition to these relations, the control circuit imposes the requirement that

$$e_c'(n)=r_cf_c'(n)+\phi_a(n)+2\phi_b(n)+\phi_c(n)+\phi_d(n) \qquad (40)$$

One now observes that the pattern of operation is repeated in the following interval $T/6<t<T/3$. Reactor *Y*1 now operates in mode *a* (gating); reactor *X*1, in mode *c* (reset throughout interval); reactors *Y*2 and *Y*3, in mode *b* (postfiring reset); and reactor *X*3 in mode *d* (postfiring 'set'). Reactor *X*2 is saturated throughout the interval.

One further observes that, following the interval in which it gates, each reactor remains saturated for one interval; and that, in the succeeding five intervals, its flux varies according to modes *b*, *c*, *b*, *d*, and *a* in that order. During this period, the reactor undergoes a full cycle of operation; and the net change in its flux is zero. Consequently, the sum of the average rates of flux change is zero; or,

$$0=\phi_a(n)+\phi_d(n-1)+\phi_b(n-2)+\phi_c(n-3)+\phi_d(n-4) \qquad (41)$$

This equation brings out the memory, or time delay, associated with the circuit. Together with equations 34 through 40, it forms a complete set of describing equations, which account for both the steady state and the dynamic operation of the circuit. A block diagram representation may readily be obtained from these equations, as demonstrated in Fig. 14.

To obtain the steady-state control characteristic, one simply solves the describing equations ignoring the cycle sequence number. The result is best presented in normalized form, as follows:

$$\frac{e_t}{2e_g}=\frac{1+\dfrac{r_m}{r_c}\dfrac{e_c'}{2e_g}}{1-\dfrac{r_c+r_m}{r_c+3r_m}K_f-\dfrac{2r_m}{r_c+3r_m}\dfrac{e_c'}{2e_g}} \qquad (42)$$

At minimum output, $e_t=0$, which requires immediately that the control voltage satisfy the relation

$$e_{c3}'=(e_{c3}+r_cF_o)=-\frac{r_c}{r_m}.2e_g \qquad (43)$$

On the other edge of the control characteristic, firing at the beginning of the middle third of each gate voltage swing, $e_t=2e_g$; and hence, the control voltage is given by

$$e_{c2}'=e_{c2}+r_cF_o=-\frac{r_c}{3r_m}K_f.2e_g \qquad (44)$$

Note that, if K_f is zero ($r_f=0$), maximum output is indeed reached here, since equation 44 then expresses the condition under which the control voltage does not contribute to the reset of the reactors. If K_f is not zero, actual maximum output will be reached with $e_c'=0$.

The circuit is said to be normally excited when, under minimum output conditions, the gate-supply voltage is just sufficient to carry each reactor between negative and positive saturation. One notes that under these conditions only modes *a* and *c* exist, so that both the reset and the setting action take place during one sixth of a cycle. The voltage which produces these changes in flux is then $2e_g$. Since, by definition, e_{go} is the voltage to set fully, or reset, the reactors during one-half cycle, as seen in equation 8, it follows that

$$e_{gn}=\frac{3}{2}e_{go} \qquad (45)$$

Some simplification of the transfer function is possible if one ignores K_f and assumes voltage control of the amplifier ($r_c \ll r_m$). Under these conditions, one may first obtain the incremental gain by differentiating equation 42

$$G_{inc}=\frac{\dfrac{r_m}{r_c}+\dfrac{2r_m}{r_c+3r_m}}{\left[1-\dfrac{2r_m}{r_c+3r_m}\dfrac{e_c'}{2e_g}\right]^2} \qquad (46)$$

The values of $e_c'/2e_g$ are next substituted from equations 43 and 44 to obtain the values of the incremental gain at the maximum and minimum output points. These are

$$\frac{r_m}{r_c}\left(1+\frac{2r_c}{r_c+3r_m}\right) \text{ and } \frac{r_m}{r_c}\Big/\left(1+\frac{2r_c}{r_c+3r_m}\right)$$

respectively. If r_m is reasonably large compared to r_c, both values reduce to r_m/r_c, which is also the average gain. In other words, it appears that under these conditions the deviations from linearity are negligible, which implies that the reset achieved in the nonlinear mode *b* is negligible compared to the reset achieved in mode *c*. One, therefore, ignores the nonlinear portion of the block diagram and arrives at a simple transfer function

$$\frac{\Delta e_t(Z)}{\Delta e_c(Z)}=\frac{\dfrac{r_m}{r_c}Z^3}{1+\left(1-\dfrac{r_m}{r_c}\right)Z^3} \qquad (47)$$

One now recognizes that

$$Z^3=e^{-\frac{T}{2}s}\sim\frac{1}{1+\dfrac{T}{2}s} \qquad (48)$$

where the approximation holds for slowly varying signal frequencies. Equation 47 then simplifies to read

$$\frac{\Delta e_t(s)}{\Delta e_c(s)}=\frac{r_m}{r_c}\cdot\frac{1}{1+\dfrac{r_m}{r_c}\dfrac{T}{2}s} \qquad (49)$$

References

1. ANALYSIS AND DESIGN PROCEDURE FOR A TWO-STAGE MAGNETIC AMPLIFIER, T. Knutrud. *M.S. Thesis*, Massachusetts Institute of Technology, Cambridge, Mass., June 1956.

2. MAGNETIC AMPLIFIER ANALYSIS AND APPLICATIONS USING BLOCK-DIAGRAM TECHNIQUES, R. M. Hubbard. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, Nov. 1957, pp. 578–88.

3. A MATHEMATICAL MODEL FOR A MAGNETIC-AMPLIFIER REACTOR CORE, H. H. Woodson. *Ibid.*, see pp. 872–78 of this issue.

4. MAGNETIC AMPLIFIER ANALYSIS USING A RESISTIVE CORE MODEL, S. Frankenthal. *M.S. Thesis*, Massachusetts Institute of Technology, Jan. 1958.

5. ANALYSIS OF MAGNETIC AMPLIFIERS BY THE USE OF DIFFERENCE EQUATIONS, P. R. Johannessen. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 73, 1954 (Jan. 1955 section), pp. 700–11.

6. MAGNETIC AMPLIFIER ANALYSIS USING A GENERALIZED MODEL FOR THE SATURABLE REACTOR CORE, H. H. Woodson. *WADC Technical Note no. 56-193*, Servomechanisms Laboratory, Massachusetts Institute of Technology, May 1956.

7. TWO-STEP SWITCHING OF SATURABLE REACTORS, B. Lovell. *AIEE Transactions*, see pp. 895–900 of this issue.

◆

Discussion

G. C. Feth (General Electric Company, Schenectady, N.Y.): In this paper the question is raised whether the resistive model can be used to predict performance characteristics of relatively complicated magnetic amplifiers *in terms of the reactor core materials*. However, the ensuing treatment of the magnetic behavior of the core material appears to be considerably oversimplified; some of the objections are outlined below:

1. The relation of rate of change of core flux $d\phi/dt$ to the magnetic intensity is approximated as $d\phi/dt$ proportional to $(H-H_c)^a$, in which a is fixed as 2. In similar work reported in reference 1 the corresponding value of the exponent was found to vary between approximately 3 and 5 for four sets of cores of this same alloy. Thus, one would question whether the exponent 2 really has any generality.

2. The form function $g(r)$ used with the assumption of constant dr/dt at constant applied field is assumed sinusoidal. Used in this way $g(r)$ is essentially the form of the functional relationship between $d\phi/dt$ and ϕ. However, a sinusoidal relationship does not appear to fit well the observed relationship between $d\phi/dt$ and ϕ as reported in Fig. 6 of reference 2; the observed peak ϕ' occurs more nearly at $\phi=\phi_s/2$ rather than at $\phi=0$ as in the sine approximation, and the form is really not much like a sinusoid, its observed inflection point occurring at $\phi\cong0$, $d\phi/dt\cong(1/3)[d\phi/dt]_{\text{peak}}$ rather than at $\phi=-\phi s$, $d\phi/dt=0$ as in the approximation.

3. This whole development, which appears intended to relate core behavior to ferromagnetic theory, is then discarded in favor of the simplification that $g(r)$ is constant and therefore $d\phi/dt$ is independent of ϕ when the core is unsaturated.

In view of these simplifications it is not surprising that, as concluded in the paper, when there is much latitude in the variation of reset rates, $d\phi/dt$ is encountered in a given circuit the accuracy of predictions based on this treatment of core behavior is affected noticeably. In fact, when considering circuits with sinusoidal excitation voltage or with time-varying control voltage one would expect the wide variation of reset rates to be a major deterrent to obtaining accurate quantitative information with this method.

REFERENCES

1. CORE-RESET FUNCTIONS IN MAGNETIC-AMPLIFIER ANALYSIS PART I—A CORE-RESET FUNCTION, G. C. Feth. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, Sept. 1958, pp, 503–09.

2. DYNAMIC CORE BEHAVIOR AND MAGNETIC AMPLIFIER PERFORMANCE, L. A. Finzi, D. L. Critchlow. *Ibid.*, vol. 76, May 1957, pp. 229–40.

Elimination of Coupling Problems in Multistage Magnetic Amplifiers

WILLIAM MCMURRAY
ASSOCIATE MEMBER AIEE

WHEN TWO or more magnetic amplifiers are cascaded, it is generally found that the over-all characteristic cannot be accurately predicted from the individual stages. The cascaded gain is less than the product of the separate gains. There may be abrupt changes in the slope of the cascaded control characteristic, and sometimes "snap-action" occurs. The over-all transfer function has a time constant lag term for each stage and perhaps additional time constants introduced by the interstage coupling network, for example, where filtering is employed.

While satisfactory performance can, in most cases, be achieved with conventional coupling methods by judicious design, the choice of the coupling circuit and its parameters often requires considerable experimentation. This increases the time required to develop a new multistage magnetic amplifier. Moreover, the performance will be less than the optimum, attained by employing a coupling method which eliminates the basic problems.

The source of the trouble is the prefiring voltage induced in the control windings of the second stage, which disturbs the resetting of the first stage. Also, the output voltage waveform of the first stage may not be ideal for efficient control of the second stage, particularly when the amplifier is push-pull.

The Synchronous Switch Coupling Method

The coupling problem can be solved by the circuit arrangements shown in Figs. 1, 3, and 9. Each gate winding of the first stage is connected in series with the control winding of the oppositely-phased reactor of the second stage. A rectifier-type synchronous switch effectively couples the stages during the gating half cycle of stage 1 (the resetting half cycle of stage 2) and decouples the stages during the resetting half cycle of stage 1 (the gating half cycle of stage 2). The decoupling action is effected by increasing the gate-limiter voltage of stage 1, thus preventing the pefiring voltage induced in the control winding of stage 2 from disturbing the resetting of stage 1.

The use of synchronous switches to improve the performance of magnetic amplifiers has been reported by a number of authors.[1-7] In particular, Voice[1] describes the application of switched diodes to coupling half-wave half-cycle response magnetic amplifiers. The basic building-block of these circuits, and of the full-wave circuits analyzed in the present paper, is the 2-stage half-wave amplifier shown in Fig. 1.

Operation of the arrangement is easily explained, with reference to Fig. 2. During the gating half-cycle of stage 1 (voltage polarities as indicated in Fig. 1) the switch diode conducts and the switch voltage e_{S1} appears across the current-limiting resistance R_{S1}. Thus, point B is effectively connected to point A (Fig. 1) and stage 1 is set by the voltage e_1 applied to its gate winding. It is assumed that the prefiring gate current required by stage 1 is much less than the control current required to reset stage 2:

$$\frac{H_{C1}l_{Fe1}}{N_{G1}} \ll \frac{H_{C2}l_{Fe2}}{N_{C2}} \tag{1}$$

When stage 1 saturates or "fires" at some angle α_1, the voltage e_1 is applied to reset stage 2 at its gate-limited rate. The necessary conditions for gate-limited reset of stage 2 are:

$$e_1 - \frac{H_{C2}l_{Fe2}}{N_{C2}}(R_{G1}+R_{C2}) > e_2\frac{N_{C2}}{N_{G2}} \tag{2}$$

$$\frac{e_{S1}}{R_{S1}} > \frac{H_{C2}l_{Fe2}}{N_{C2}} \tag{3}$$

Then in the next half cycle, stage 2 will fire at an angle α_2, where

$$\alpha_2 = \pi - \alpha_1 \tag{4}$$

The voltage polarities now being reversed, the switch diode will block, effectively connecting point B to point C (Fig. 1). The prefiring voltage e_2N_{C2}/N_{G2} induced in the control winding of stage 2 will not limit the resetting of stage 1, provided that

$$e_{S1} > e_2\frac{N_{C2}}{N_{G2}} \tag{5}$$

The winding voltage of stage 1 during reset is determined by the applied control voltage, the control circuit resistance R_{C1}, the dynamic core properties and the gate-limiter voltage. The conditions differ from the single-stage case only in the extension of the gate-limiter voltage, particularly after stage 2 fires at the angle α_2. This will have very small effect if half-sine wave voltage control is used, but may have appreciable effect when current-source control is employed. In any case, the effect will not be detrimental to gain or response.

Under steady-state conditions, the gate volt-second area induced during reset of stage 1 will equal the integral of the voltage e_1 up to the previously assumed firing angle α_1. In Fig. 2, an arbitrary but typical shape of the resetting voltage is sketched.

It is seen from equation 4 that the output of stage 2 decreases as the output of stage 1 increases. The over-all control characteristic is similar in shape to the inverted characteristic of stage 1 alone. A change in the output of stage 1 results in a corresponding change in the output of stage 2 during the next half-cycle. Thus stage 2, and any subsequent stages, has half-cycle response. A multistage amplifier exhibits the control circuit time constant of the first stage, plus a delay time of one half-cycle times the number of stages.

Synchronous Switch Coupling of Full-Wave Amplifiers

The operation of the coupling method with full-wave circuits is generally similar to the half-wave case, since the interstage connections are essentially half-wave. However, when the final stage is a center-tap circuit, a complication in the mode of operation can occur as a result of the increase in the gate-limiter voltage of the resetting reactor after the gating reactor saturates. A 2-stage full-wave d-c single-ended amplifier is shown in Fig. 3 and is analyzed in the Appendix. Both stages employ the center-tap circuit with the addition of a synchronous switch.

The switch in stage 2 is not necessary for coupling the stages but affords some improvement in performance by increasing the gate-limiter voltage of stage 2. This effect has been termed "firing angle amplification"[1] or "pulse stretching"[3] and

Vol.
Pt. 1
195

Paper 59-1053, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-IRE Nonlinear Magnetics and Magnetic Amplifiers Conference, Washington, D. C., September 23-25, 1959. Manuscript submitted June 18, 1959; made available for printing July 16, 1959.

WILLIAM MCMURRAY is with General Electric Company, Schenectady, N. Y.

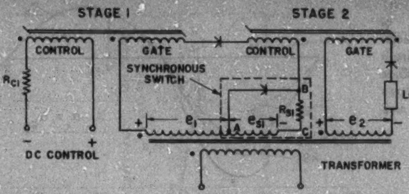


Fig. 1 (above). Two-stage half-wave magnetic amplifier with synchronous switch coupling

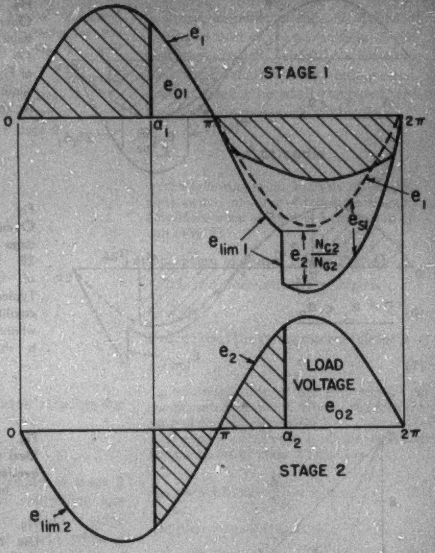


Fig. 2 (right). Operation of half-wave magnetic amplifiers with synchronous switch coupling. Gate winding voltage is shaded area

enables full-range control of stage 2 to be obtained with less than full-range control of stage 1. The stage 2 switch current is inductively filtered, enabling the unimpeded passage of the "rectangular blocks" of gate current required by the inductive load. An almost similar result could be achieved by increasing the magnitude of unfiltered switch current, but at the expense of greater switch dissipation.

Control and bias of stage 1 are obtained by control windings wound around both cores in the normal manner, with mode of operation and time constant identical to that of a conventional full-wave circuit, except as influenced by the modified gate-limiter voltage. Typical voltage wave forms are sketched in Figs. 4 and 5. The stage 1 resetting voltage is shown for the ideal low-impedance control circuit case.[8]

Ideal normalized transfer characteristics, relating the per unit output voltage of stage 2 to the per unit output voltage of stage 1, are shown in Fig. 6 for several typical values of circuit parameters. Use of the transfer characteristics to predict the over-all control characteristic from the characteristic of the first stage is illustrated in Fig. 7. It is seen that the effect of the center-tap connection of stage 2 is to increase the over-all gain when $\alpha_1 > \alpha_2$ and to decrease the over-all gain when $\alpha_1 < \alpha_2$. Since the gain of stage 1 is generally highest in the lower part of its output range ($\alpha_1 > 90$ degrees), the cascaded characteristic accentuates the nonlinearity of the first stage. However, the low gain region is often avoided in normal operation and the increase in the maximum gain is welcome. The modification of the characteristic is dependent upon the ability of stage 1 to reset stage 2 at a rate greater than the minimum requirements, that is, the parameter k must be greater than unity. Without a switch in stage 2, the condition $k=2$, $s_2=0$ (Fig. 6) represents the limiting transfer characteristic obtainable.

When a switch is employed in stage 2 the gain can be further increased and the nonlinearity reduced. The high gain region of stage 1 can be used to control stage 2 over its full range (see Fig. 7).

A typical over-all control characteristic, without a switch in stage 2, is shown in Fig. 8. The coupling parameter k is approximately 1.3. The first stage employs Supermalloy cores, and the second stage Deltamax cores. The resemblance to an inverted single-stage characteristic is obvious (no bias is used).

Full-Wave Push-Pull Circuit

Two similar single-ended circuits can be combined in a push-pull connection of the stage 1 control windings and of the stage 2 load, but the interstage push-pull connection of conventional circuits is avoided. The synchronous switch circuit is common to both sides, as shown in Fig. 9. The coupling of each side operates independently in the manner described for the single-ended circuit of Fig. 3.

Typical cascaded control characteristics, with a Supermalloy core first stage and a Deltamax core second stage, are shown in Fig. 10. Curve A is obtained without a switch in stage 2 ($k=2$, $s_2=0$), and curve B illustrates the effect of adding a switch in stage 2 ($k=2$, $s_2=1/2$). With the switch, the linear range is extended to the saturation level. The load is a torque motor with twin windings

A previous design using the same cores, but with conventional dummy load coupling, had insufficient dynamic gain to meet the specifications of the application, and poor linearity. Attempts to increase the gain produced bistable regions in the characteristic. Linearity could be improved only at the expense of gain. Both problems were solved by the use of synchronous switch coupling.

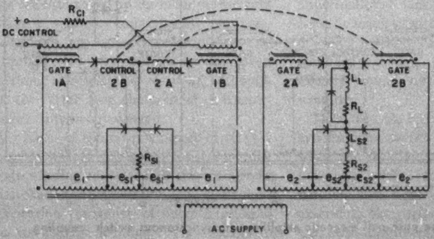


Fig. 3. Two-stage full-wave magnetic amplifier with synchronous switch coupling

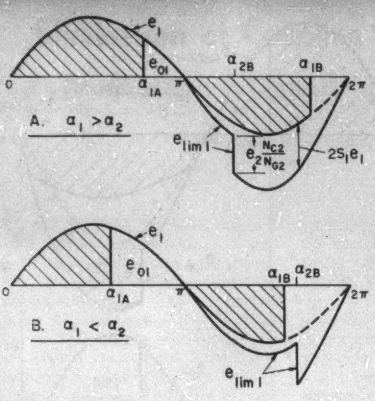


Fig. 4 (left). Operation of stage 1 reactor 1A in circuit of Fig. 3. Gate winding voltage is shaded area

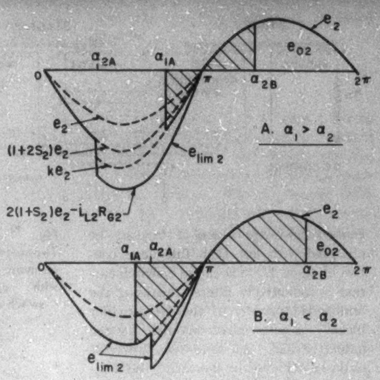


Fig. 5 (right). Operation of stage 2 reactor 2B in circuit of Fig. 3. Typical Case II conditions. Gate winding voltage is shaded area

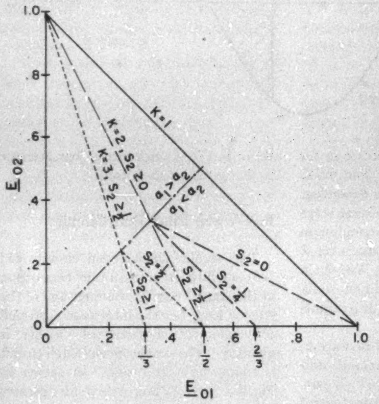


Fig. 6 (left). Ideal normalized transfer characteristics

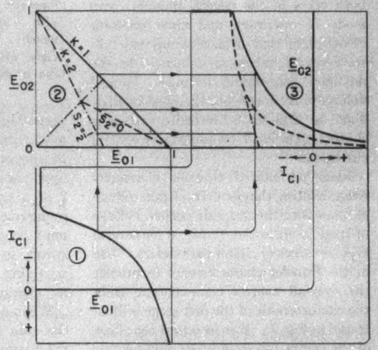


Fig. 7 (right). Use of transfer characteristic to predict control characteristic of cascaded amplifiers

1 — Normalized control characteristic of stage 1

2 — Normalized transfer characteristic

3 — Normalized control characteristic of cascaded amplifiers

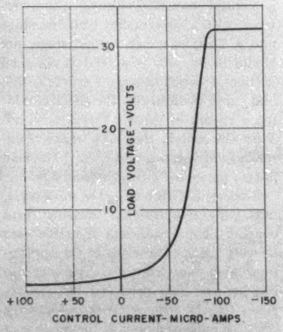


Fig. 8. Typical cascaded control characteristic for circuit of Fig. 3

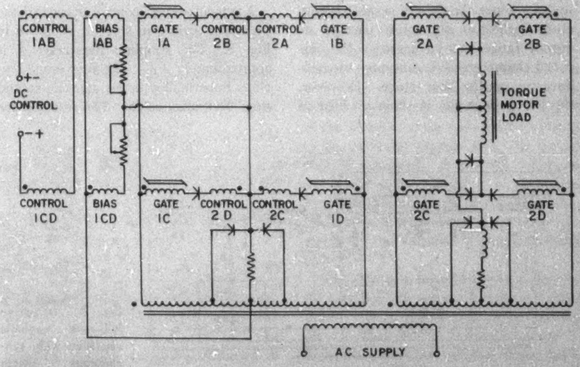

CONTROL 1AB
BIAS 1AB
GATE 1A
CONTROL 2B
CONTROL 2A
GATE 1B
GATE 2A
GATE 2B
DC CONTROL
TORQUE MOTOR LOAD
CONTROL 1CD
BIAS 1CD
GATE 1C
CONTROL 2D
CONTROL 2C
GATE 1D
GATE 2C
GATE 2D
AC SUPPLY

Fig. 9. Two-stage push-pull magnetic amplifier with synchronous switch coupling

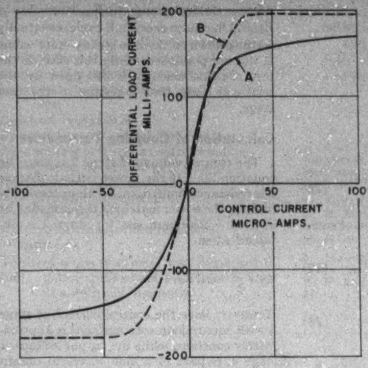


Fig. 10. Typical cascaded control characteristics for circuit of Fig. 9

A—Without synchronous switch in stage 2
B—With synchronous switch in stage 2

Conclusions

The properties of the synchronous switch coupling method can be summarized by listing the advantages and disadvantages of the method.

ADVANTAGES

1. Stage 2 is prevented from influencing the resetting of stage 1. The tendency toward nonlinearities which occurs with conventional methods (such as dummy load coupling of push-pull amplifiers) is eliminated.
2. No interstage time constant is introduced. Stage 2 has half-cycle response.
3. Since the stage 2 reactors are voltage reset, and effectively slaves of stage 1, the effects of drift and mismatch of the stage 2 cores are reduced to second-order importance. Only the drift and matching of stage 1 cores are critical.
4. No bias is required for stage 2.
5. The synchronous switch circuit can serve as a d-c bias source for stage 1, as shown in Fig. 9.
6. The theory of coupling is simple and performance is predictable.
7. The method is equally applicable to half-wave, full-wave, single-ended, or push-pull circuits.
8. Problems caused by inductive loading are eliminated by the mode of resetting stage 2. Silicon commutating rectifiers, with high forward voltage drop, cause no trouble. No snap-action occurs, even without commutating rectifiers, although in the latter case the control characteristic is drastically altered.
9. None of the output power of stage 1 need be wasted in coupling resistors. The gate winding resistance of stage 1 can be matched approximately to the equivalent excitation resistance of stage 2 referred to its control winding. This results in minimum size and shortest response for stage 1.

DISADVANTAGES

1. The full range of output of stage 2 cannot be obtained by controlling less than the full range of the output of stage 1, except to a limited extent when a synchronous switch is also used in stage 2. This restriction, however, also applies to most other methods of coupling without interstage filtering.
2. Power is dissipated in the switch. However, an equal amount of power may normally be required to bias stage 2, and is generally a relatively small amount. If a switch is used in stage 2 its power dissipation is a definite drawback if the load is large or the source is of limited capacity.
3. The "center-leg" method of self-biasing push-pull amplifiers cannot be used with an even number of stages.
4. The method cannot be used for coupling single phase to polyphase stages, or stages operating from different supply frequencies.
5. Due to the inversion of the characteristic, the minimum load current may be rather large especially when the core material of stage 1 has a low ratio of residual to saturated flux density, such as Supermalloy. This fault can be overcome by using a switch in stage 2.
6. The method cannot be applied to standard magnetic amplifiers which are already wound with a common control winding.

It is seen that the benefits gained by synchronous switch coupling are many, while the disadvantages listed are often not applicable or not serious. The method offers a solution to the problem of extending the performance of multistage magnetic amplifiers. The shunt-coupled magnetic amplifier circuits of Hubbard[9] also have most of the advantages of synchronous switch coupling. However, they require that the gate winding resistance of stage 1 be small compared with the excitation resistance of stage 2.

Appendix

The following is an analysis of synchronous switch coupling of full-wave magnetic amplifiers when stage 2 is a center-tap circuit (Fig. 3).

Operation of Stage 1 (see Fig. 4)

The synchronous switch voltage is

$$e_{S1}=s_1e_1 \tag{6}$$

Before firing, the gating reactor voltage is

$$e_{G1}=e_1 \tag{7}$$

(The forward drop of the gate rectifier is compensated approximately by the forward drop of the switch rectifier.)

The gate limiter voltage of the resetting reactor is, before stage 2 fires at α_2

$$e_{\lim 1}=(1+2s_1)e_1-\frac{N_{C2}}{N_{G2}}e_2 \tag{8}$$

and after stage 2 fires at α_2

$$e_{\lim 1}=(1+2s_1)e_1 \tag{9}$$

From equation 8, in order to prevent stage 2 from affecting stage 1, requires $e_{\lim 1}>e_1$.

$$2S_1>\frac{N_{C2}}{N_{G2}}\frac{e_2}{e_1} \tag{10}$$

Operation of Stage 2 (see Fig. 5)

The commutating rectifier forward drop is neglected.

A. WITH SYNCHRONOUS SWITCH VOLTAGE

$$e_{S2}=s_2e_2 \tag{11}$$

Before firing, the gating reactor voltage is

$$e_{G2}=e_2 \tag{12}$$

After firing, the load voltage is

$$e_{L2}=e_2-i_{L2}R_{G2} \tag{13}$$

The gate limiter voltage of the resetting reactor is, before the gating reactor fires at α_2

$$e_{\lim 2}=(1+2s_2)e_2 \tag{14}$$

and after the gating reactor fires at α_2

$$\begin{aligned} e_{\lim 2}&=(1+2s_2)e_2+e_{L2}\\ &=2(1+s_2)e_2-i_{L2}R_{G2} \end{aligned} \tag{15}$$

B. WITHOUT SYNCHRONOUS SWITCH

Without the synchronous switch, $s_2=0$

Equations 12-15 become, respectively,

$$e_{G2}=e_2-e_f \tag{12A}$$

$$e_{L2}=e_2-e_f-i_{L2}R_{G2} \tag{13A}$$

$$e_{\lim 2}=e_2+e_f \tag{14A}$$

$$\begin{aligned} e_{\lim 2}&=e_2+e_f+e_{L2}\\ &=2e_2-i_{L2}R_{G2} \end{aligned} \tag{15A}$$

IDEAL NORMALIZED TRANSFER CHARACTERISTIC

The per unit output of stage 1 is

$$\underline{E}_{o1}=\frac{E_{o1}}{E_1}=\frac{1+\cos\alpha_1}{2} \tag{16}$$

The per unit output of stage 2 is

$$\underline{E}_{o2}=\frac{E_{o2}}{E_2}=\frac{1+\cos\alpha_2}{2} \tag{17}$$

The per unit reset of stage 2 required is

$$\frac{E_{G2}}{E_2}=1-\underline{E}_{o2}=\frac{1-\cos\alpha_2}{2} \tag{18}$$

Define the parameter k after stage 1 fires at α_1 as

$$k=\frac{e_{C2}}{e_2}\frac{N_{G2}}{N_{C2}} \tag{19}$$

where

$$e_{C2}=e_{o1}-i_{C2}(R_{G1}+R_{C2}) \tag{20}$$

In order to control stage 2 over its full range we require $k \geq 1$. The maximum value of k is gate limited to

$$k_{\max}=\frac{e_{\lim 2}}{e_2} \tag{21}$$

CASE I

$1<k<1+2s_2$

No discontinuity will be caused by the center-tap circuit of stage 2. The per unit reset of stage 2 is $k\underline{E}_{o1}$

Equating this to the required value of equation 18, one obtains

$$\underline{E}_{o2}=1-k\underline{E}_{o1} \tag{22}$$

Thus

$\underline{E}_{o2}=0(\alpha_2=\pi)$

when

$$\underline{E}_{o1}=\frac{1}{k}\left(\cos\alpha_1=\frac{2}{k}-1\right)$$

CASE II

This occurs only with the center-tap circuit, where the gate-limiter voltage changes at α_2.

(a). When

$$\alpha_1>\alpha_2,\ 1+2s_2<k<2(1+s_2)-\frac{i_{L2}R_{G2}}{e_2}$$

Equation 22 is valid in this range.

When $\alpha_1=\alpha_2=\alpha_c$, one obtains from equation 22

$$\frac{1+\cos\alpha_c}{2}=1-k\ \frac{1+\cos\alpha_c}{2} \tag{23}$$

from which

$$-\cos\alpha_c=\frac{k-1}{k+1} \tag{24}$$

and

$$\underline{E}_{o2}=\underline{E}_{o1}=\frac{1}{1+k} \tag{25}$$

(b). When $\alpha_1<\alpha_2$

$k=1+2s_2$ from α_1 to α_2

$1+2s_2<k<2(1+s_2)-\dfrac{i_{L2}R_{G2}}{e_2}$ from α_2 to π

The per unit reset of stage 2 is

$$(1+2s_2)\frac{(1+\cos\alpha_1)}{2}+(k-1-2s_2)\frac{(1+\cos\alpha_2)}{2} \tag{26}$$

Equating this to the required value of equation 18 one obtains

$$1-\underline{E}_{o2}=(1+2s_2)\underline{E}_{o1}+(k-1-2s_2)\underline{E}_{o2} \tag{27}$$

from which

$$\underline{E}_{o2}=\frac{1-(1+2s_2)\underline{E}_{o1}}{k-2s_2} \tag{28}$$

Thus

$\underline{E}_{o2}=0(\alpha_2=\pi)$

when

$$\underline{E}_{o1}=\frac{1}{1+2s_2}\left(\cos\alpha_1=\frac{2}{1+2s_2}-1\right)$$

Typical conditions of Case II are shown in Fig. 5. From the foregoing information, the normalized transfer characteristic may be constructed for particular values of s_2 and k. Examples are shown in Fig. 6, and the application of the transfer characteristic to predict the control characteristic of the cascaded amplifiers is illustrated in Fig. 7.

Without the synchronous switch in stage 2, $(s_2=0)$ the equations are simplified:

CASE I

$k=1$

$$\underline{E}_{o2}=1-\underline{E}_{o1} \tag{22A}$$

CASE II

(a). When

$$\alpha_1>\alpha_2,\ 1<k<2-\frac{i_{L2}R_{G2}}{e_2}$$

$$\underline{E}_{o2}=1-k\underline{E}_{o1} \tag{22}$$

(b). When $\alpha_1<\alpha_2$

$k=1$ from α_1 to α_2

$1<k<2-\dfrac{i_{L2}R_{G2}}{e_2}$ from α_2 to π

$$\underline{E}_{o2}=\frac{1-\underline{E}_{o1}}{k} \tag{28A}$$

A deviation from half-cycle response occurs under Case II conditions when $\alpha_1<\alpha_2$. This is due to the change in gate-limiter voltage of the center-tap circuit when the gating core fires. Since the latter firing is predetermined during the previous half-cycle, and affects the resetting during the present half-cycle which determines the output during the next half cycle, a channel for the hold-over of information for more than one half-cycle is established.

For transient conditions, equation 27 becomes

$$\begin{aligned}1-\underline{E}_{o2}[(n+1)\text{th half-cycle}]\\ =(1+2s_2)\underline{E}_{o1}[n\text{th half-cycle}]+\\ (k-1-2s_2)\underline{E}_{o2}[n\text{th half-cycle}]\end{aligned} \tag{29}$$

After a sudden change in $\underline{E}_{o1}$, the value of $\underline{E}_{o2}$ will fluctuate every half-cycle, eventually settling down to the new steady-state value. It was not possible to verify this effect in the test circuits, since it is swamped by the relatively long time constant of the first stage.

Calculation of Coupling Parameters

The control voltage of stage 2 is given by equation 20. If the power output of stage 1 is resistance limited, rather than thermally limited, the power matching theorem may be applied. Minimum size for stage 1 is obtained when

$$e_{C2}=\frac{e_{o1}}{2}=i_{C2}(R_{G1}+R_{C2})=ke_2\frac{N_{C2}}{N_{G2}} \tag{30}$$

However, since the control current of stage 2 with square-loop core material is approximately constant, while the output voltage of stage 1 is part of a sine wave, to obtain good control near the ends of the range ($\alpha=$ 0+ or $\pi-$) it is desired that

$$E_1>2I_{C2}(R_{G1}+R_{C2}) \tag{31}$$

where

$$I_{C2}=\frac{H_{C2}l_{Fe2}}{N_{C2}} \tag{32}$$

As a design guide, one may choose

$$N_{C2}\approx\frac{E_1}{E_2}\frac{N_{G2}}{2} \tag{33}$$

and check that condition 31 is satisfied. An approximate average value for the parameter k may now be calculated.

$$k=\frac{E_1-I_{C2}(R_{G1}+R_{C2})}{E_2}\cdot\frac{N_{G2}}{N_{C2}} \tag{34}$$

Whether or not k will be gate-limited to a value smaller than that given by equation 34 is determined by the upper limit of the inequalities given for the conditions of Case I or Case II operation.

Nomenclature

e = gate supply voltage, absolute instantaneous value
E = gate supply voltage, half-cycle average value
e_C = control winding voltage, absolute instantaneous value
e_f = forward voltage drop of diode
e_G = gate winding voltage, absolute instantaneous value
E_G = gate winding voltage, half-cycle average value
e_L = load voltage, instantaneous value
$e_{\lim}$ = gate limiter voltage: maximum magnitude of gate winding voltage during reset half-cycle, instantaneous value
e_o = gross output voltage, absolute instantaneous value
E_o = gross output voltage, half-cycle average value
$\underline{E}_o$ = per unit output voltage E_o/E
e_s = synchronous switch voltage, absolute instantaneous value
H_C = coercive force of core material, ampere-turns per cm (centimeter)

i_C = control current, absolute instantaneous value
I_C = control current, half-cycle average value
i_L = load current, instantaneous value
k = parameter defined by equation 19
l_{Fe} = mean magnetic circuit length of core, cm
L_L = load inductance
L_S = synchronous switch inductance
N_C = number of turns in control winding
N_G = number of turns in gate winding
R_C = conrol circuit resistance
R_G = gate winding resistance
R_L = load resistance
R_S = synchronous switch resistance
s = ratio e_S/e
α = firing angle measured from beginning of gating half cycle, electrical radians at a-c supply frequency
π = 3.14...
Subscripts 1,2 designate amplifier stages 1 and 2
A,B,C,D designate reactors in full-wave circuits

References

1. FAST RESPONSE MULTI-STAGE MAGNETIC AMPLIFIERS, C. C. Voice. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 12, 1956, pp. 411–24.

2. THE PULSE STRETCH COUPLING CIRCUIT, H. W. Patton. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, Sept. 1956, pp. 377–9.

3. MAGNETIC-AMPLIFIER CIRCUITS WITH FULL-WAVE OUTPUT AND HALF-WAVE CONTROL SIGNALS, H. W. Lord. *Ibid.*, vol. 73, July 1954, pp. 265–70.

4. RAPID RESPONSE SELF-SATURATING MAGNETIC AMPLIFIER, B. D. Bedford. *U. S. Patent no. 2,733,306*, Jan. 31, 1956.

5. FULL-WAVE REVERSIBLE POLARITY HALF-CYCLE RESPONSE MAGNETIC AMPLIFIERS, C. B. House. *Report no. 4541*, Naval Research Laboratory, Washington, D. C., June 20, 1955.

6. CASCADE CONNECTIONS OF FAST RESPONSE MAGNETIC AMPLIFIERS, Yoshifumi Sakurai. *Technology Reports*, Osaka University, Osaka, Japan, vol. 6, no. 190, 1956.

7. HIGH-EFFICIENCY PUSH-PULL MAGNETIC AMPLIFIERS WITH TRANSISTORS AS SWITCHED RECTIFIERS, A. G. Milnes. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, July 1958, pp. 327–30.

8. THEORY OF MAGNETIC AMPLIFIERS WITH SQUARE-LOOP CORE MATERIALS, H. F. Storm. *Ibid.*, vol. 72, Nov. 1953, pp. 629–40.

9. SHUNT-COUPLED MAGNETIC-AMPLIFIER CIRCUITS, R. M. Hubbard. *Ibid.*, vol. 78, May 1959, pp. 124–31.

Miniature High-Frequency Magnetic Amplifiers for Magnetic Memory Sensing

G. R. BRIGGS
NONMEMBER AIEE

K. C. HU
ASSOCIATE MEMBER AIEE

AMPLIFIERS of the magnetic modulator type have been found useful for amplifying signals as small as 10^{-18} to 10^{-17} watts.[1,2] The most common design utilizes balanced cores and a flux-coupled output circuit separate from the excitation circuit. The output is the difference output of the cores and is predominantly second harmonic in content. The zero drift is small because core unbalance results only in generation of a fundamental output component which can be supressed by filtering,[3] and the elimination of "series-connected" loads in the excitation circuit reduces drifts arising from excitation amplitude fluctuations.[4] Parallel connection of windings which introduces problems of d-c resistance balancing, and the use of output rectifiers which results in drifts arising from d-c feedback, are also avoided to reduce zero drift further.

In optimizing the circuit in these ways to obtain drift-free operation, the amplifier becomes inherently simple in design. This simplicity implies that this type of amplifier should exhibit fast response. Lufcy and his co-workers pointed out some time ago[5] that as few cores and as few paths for d-c flow as possible are desirable to obtain fast response, because it is necessary to keep the effective flux and the number of current paths having significant L/R (inductance-resistance) time constants to a minimum. The balanced-core magnetic modulator satisfies these requirements well and therefore, should exhibit a minimum response time nearly as small as the minimum possible of one half of the excitation period. Because of this potential speed possibility, it was felt desirable to investigate this circuit as an output-signal detector for a magnetic memory. This application requires the use of excitation frequencies greater than 100 kc, but very little has been published concerning response for frequencies greater than 400 cps (cycles per second). Kletsky[6] has recently reported gain measurements at 6,000 cps, but he did not report measurements of response time. The lack of high-frequency data has prompted the present investigation in the frequency range of 100 kc to 4 mc (megacycles).

At such high frequencies of excitation it might be expected that useful voltage and power gains could be obtained with small cores having small fluxes and wound with few turns of wire. The measurements to be reported were made on amplifiers constructed of ferrite and metal tape cores approximately 0.050 inch in diameter, wound with single-turn control windings. To keep the circuit simple, no filtering to remove the fundamental component of excitation frequency in the output was provided. This limitation, in conjunction with the effects of core unbalance caused by high-frequency heating and the use of an asymmetric mode of excitation to obtain greater gain and gain stability, reduced the zero stability in comparison with low-frequency designs. A signal power sensitivity of about 10^{-10} watts can be obtained. This moderate power sensitivity is accompanied by a high voltage senstivity (noise is less than 0.1 μv (microvolts) equivalent at the input) and a response time less than 10 μsec (microseconds), making the device attractive for applications involving low level voltage signals of moderate current.

General Considerations

CIRCUIT

A schematic diagram of the amplifier is shown in Fig. 1. The circuit shown is similar to one described by Wennerberg.[7] In this circuit a diode shunted by a resistance, R_s, is connected in series with the excitation winding to obtain an asymmetrical excitation current. This will later be shown to result in more gain and less dependence of gain on excitation amplitude as well as faster response. The output and control windings have the same winding sense, with the result that output appears across both. The input circuit, therefore, loads the output. if the source impedance, R_c in Fig. 1, is sufficiently high, the output loading is negligible, but if R_c is small, the output

---

Paper 59-1042, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-IRE Nonlinear Magnetics and Magnetic Amplifiers Conference, Washington, D. C., September 23–25, 1959. Manuscript submitted June 19, 1959; made available for printing August 4, 1959.

G. R. BRIGGS and K. C. HU are with Radio Corporation of America, Princeton, N. J.

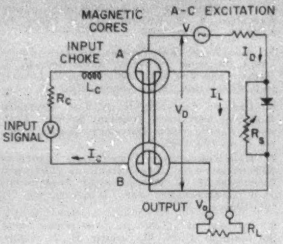


Fig. 1. Simplified schematic diagram of balanced-core magnetic amplifier

may become too heavily loaded and isolation between the source and output is then required. This can be either resistive or inductive or a combination of the two. Inductive isolation is usually chosen to preserve the maximum possible gain. This inductance is shown as L_c in Fig. 1. The time constant L_c/R_c will mainly determine the response time, as will later be shown, when this time is large compared to one-half period of the excitation and if the cores are operated in a saturating mode.

The output load is a resistance, R_L. The use of rectifying circuits here is avoided to eliminate circuit analysis complications introduced by the resulting d-c feedback and to obtain fastest response. A resistive load is a close approximation to the load actually used in the memory sensing application for which the amplifier was developed.

SYMMETRICAL EXCITATION

The symmetrical mode of excitation obtained when $R_s=0$ is the common one for low-frequency operation and will be described for later comparison with other modes (see Fig. 2). For this mode the excitation amplitude is made sufficiently large to switch both cores alternately from one saturation level to the other, and both cores switch equally in flux. Assuming the hysteresis loop of each core is the same, their outputs will exactly cancel if the control current, I_c, is zero, as indicated in Fig. 2(B). However, if a control current is applied, the output waveforms of the cores will become asymmetrical and a net even-harmonic waveform will appear at the output. The asymmetry is a result of differences in switching delay and differences in switching speed between the cores caused by the control-current bias. Note that the voltage-time integral of the output of each core remains the same whether I_c is zero or not.

If a step of control current is applied to the amplifier the full difference output will be obtained within one-half cycle of the excitation frequency. The only additional limitation on response is determined by those control circuit parameters which act to limit the rate of change of I_c.

NONSATURATING EXCITATION

If the amplitude of the symmetrical excitation is increased beyond that required to saturate the cores, the control sensitivity decreases, because the cores tend to switch more nearly at the same rate and in the same time, producing more complete cancellation at the output. This is clearly shown by data presented in the paper already mentioned by Kletsky.[6] His results indicate that this effect becomes more important at higher frequencies of excitation. To avoid this decrease in gain with increase in excitation drive, it is best to drive the cores just to saturation. Greater gain can even be obtained by operating with a drive insufficient to saturate, but if this is done the possibility of unstable operation arises. This is characterized by an excessive dependence of gain upon drive amplitude[1] and by a large response time accompanied by considerable nonlinearity between the output and input, even for small signals. How this can happen will be explained with the aid of Fig. 3. Assuming that I_d is insufficient to saturate the cores in either of their two saturation levels, the flux excursions which occur upon application of I_c can be quite complex, as indicated by the flux excursion paths shown drawn dashed in Fig. 3(A). Assume that a step-function I_c is applied. This causes the fluxes, starting at the points indicated by *a* in the figure, to follow complex minor loop excursions on succeeding excitation cycles until the equilibrium points *d* are reached. This may take a number of cycles dependent upon the excitation and signal amplitudes. This "creeping" of unsaturating minor loops is especially marked with square-loop materials[8] and results in an output response of the form indicated in Fig. 3(B). Referring to this figure, before the control current step is applied the cores operate with symmetrical equilibrium minor loops and the net output is zero. When the step is applied considerable unbalance immediately occurs and output is obtained on the next excitation half-cycle following the step. On following cycles, however, the output does not remain constant, but continues to change until a new minor loop equilibrium condition has been re-established. This results in a long response time which is both excitation- and signal-amplitude dependent. In addition, the total amount of creep is also dependent upon these parameters and causes the final output to be nonlinear with respect to the input.

NONSYMMETRICAL EXCITATION

To avoid the aforementioned difficulties and yet obtain high gain, it is possible to use asymmetrical excitation which saturates the cores in one direction only (see Fig. 4). The saturating polarity of drive insures operation without creeping of the operating point, whereas the nonsaturating drive of opposite polarity presets the cores to sensitive operating points on their hysteresis loops. This mode of operation is common for high-level magnetic amplifiers. Wennerberg has described a mode of operation similar to this for amplifying signals in the millivolt range.[7] As indicated in Fig. 4(A), the flux change in each core is now no longer the same when I_c is applied. This difference in flux, accompanied by differences in switching rate and switching delay, now determines the net output as shown in Fig. 4(B). Note that the output is no longer even harmonic, but contains the fundamental and odd harmonics

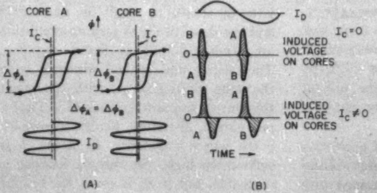


Fig. 2 (left). Operation with symmetrical, saturating excitation

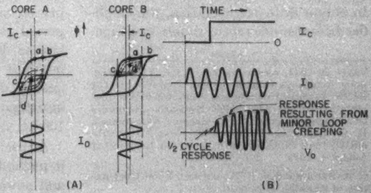


Fig. 3. Operation with symmetrical, nonsaturating excitation

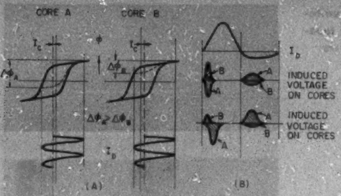


Fig. 4 (left). Operation with assymmetrical excitation

of the excitation frequency as well. The gain can be high, because if the cores are preset by the drive to a flux level of about one half of the total switching flux, the operating points are placed in regions of high incremental permeability. The gain also remains high at high excitation frequency in contrast to the symmetrically-driven amplifier, because the gain now is not as strongly dependent upon differences in switching rate and delay of the cores, effects which tend to diminish as the frequency increases.

The dependence of gain upon excitation amplitude with asymmetrical drive is as follows. Increase of the saturating polarity of drive current causes the flux differences set by I_c to be switched at a faster rate, generating greater peak voltages which increase the voltage gain and, to a lesser extent, the power gain. The effect of increasing the presetting polarity of drive current is more complex. If the cores are preset less than halfway by the drive, increased drive increases the gain, whereas if the cores are set more than halfway, increased drive decreases the gain. By setting more than halfway and by maintaining a constant ratio between saturating and setting current peak values it is possible to make the opposing gain-drive dependences cancel each other over a considerable range of excitation-source amplitude (see Fig. 5). The shunted diode shown in Fig. 1 accomplishes this; R_s is adjusted until the best operation is obtained.

The previously mentioned advantages of asymmetric excitation are offset to some extent by a greater dependence of zero balance on core temperature with this excitation mode. In the symmetrically-driven amplifier, variations in both core remanent flux and coercive force due to temperature changes result in the generation of a fundamental frequency output component which can be removed by filtering. In the asymmetrically driven amplifier, on the other hand, variations in the relative coercive forces of the cores due to temperature changes result in effects which are identical with those of true signals and these cause unavoidable zero drift. In memory applications, the high-gain high-frequency performance is usually found to be more important than zero stability and asymmetrical drive is therefore of most interest here.

Experimental Results

Ferrite-Core Amplifier

Ferrite cores of 0.050 inch outside diameter, 0.030 inch inside diameter, and 0.015 inch thickness were wound and tested in the circuit which is shown in Fig. 6. The ferrite is a low-drive type having a coercive force of about 0.06 oersted and exhibits a hysteresis loop of moderate squareness. The cores are designated as Radio Corporation of America type *XF-4008*. Switching time data for these cores obtained from rectangular pulse measurements are presented in Fig. 7. For switching rates in the frequency range near 1 megacycle which are of interest here, these data yield a switching coefficient, *Sw*, of 0.46 oersted-μsec.[9] The switching flux of each core is about 2.8 maxwells.

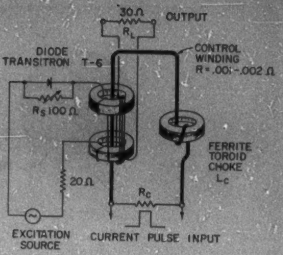


Fig. 6. Experimental test circuit for ferrite or metal core amplifier

The control winding is a single short length of no. *28* wire (see Fig. 8). The choke (not shown) is wound with even heavier wire to keep the total control circuit resistance small (0.002 ohm) to obtain a high input voltage sensitivity. The choke core is a small toroid of ferrite having low loss characteristics and low remanent flux (General Ceramics type *H*, no. *259* core). (Inductance values are given in the table associated with Fig. 13.) The control signal source for test purposes was a constant-current pulse generator. This was shunted by R_c to obtain a voltage-source circuit. An excitation winding of two turns and an output winding of four turns were used.

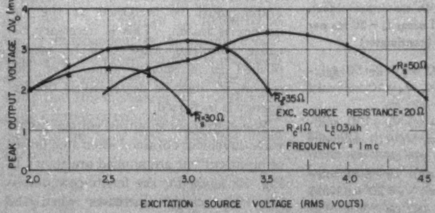


Fig. 5. Output versus excitation source voltage for different values of diode shunt resistance R_s. Data for ferrite-core amplifier

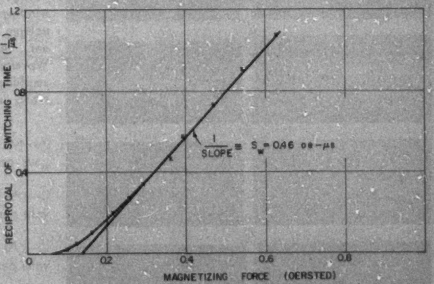


Fig. 7. Switching rate versus switching magnetic field for ferrite amplifier core

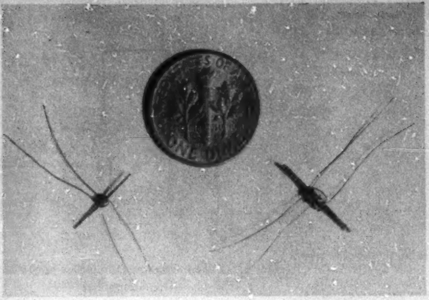


Fig. 8 (left). Ferrite-core and metal tape-core amplifiers

METAL-TAPE CORE AMPLIFIER

Metal-tape bobbin cores were also tested using the same circuit. These cores are stainless steel bobbin types wound with 22 wraps of 1/8-mil *4-79* molybdenum permalloy tape, 1/32 inch in width. The mean diameter is 0.052 inch. This diameter is considerably greater than the mean diameter (0.040 inch) of the ferrite cores, but the coercive force for the metal is less, about 0.04 oersted, so it is reasonable to expect comparable sensitivities from amplifiers constructed using the two types of cores. The switching flux of each metal core is approximately 7 maxwells. This flux is considerably larger than that of the ferrite core. With the larger flux, it is reasonable to expect a greater output and gain per turn and a slower response. However the same value of load resistance, $R_L=30$ ohms, was used in both cases, with the result that the greater output was forced to supply more load current which reduced the gain. Comparison of the amplifier gains, therefore, did not justify the use of normalization.

A photograph of the metal-core amplifier is included in Fig. 8. The larger core diameter permits the use of a heavier control conductor of lower resistance. An input resistance of 0.001 ohm could be obtained in this case.

AMPLIFIER WAVEFORMS

Typical waveforms of excitation current I_d, excitation winding voltage V_d, and output voltage V_o are shown with phase relationships preserved in Fig. 9. In all waveform photographs, time increases toward the left. These waveforms were observed for the ferrite-core amplifier with no input signal. V_o in this case is that due to core unbalance. This waveform changes in amplitude but remains relatively unchanged in form when an input signal is present. Waveforms for the metal-core amplifier are similar. Waveforms for four different excitation frequencies are presented. At low frequency, the cores in being driven to saturation switch in a small fraction of a cycle time and generate sharp spikes of voltage. At high frequency, this switching occupies a larger fraction of the cycle because the switching time remains nearly constant (constant saturating drive being assumed). For this reason also, the output voltage does not increase linearly with frequency, but at a lesser rate.

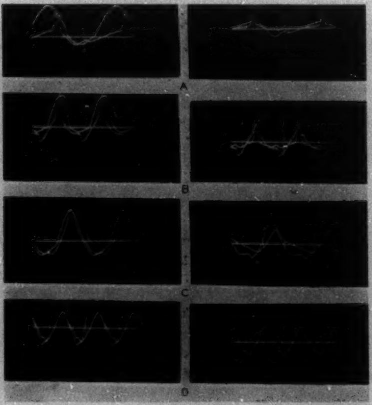


Fig. 9. Waveforms of excitation current I_d and excitation winding voltage V_d (left) and excitation winding voltage V_d and output voltage V_o (right) for four different excitation frequencies. Waveforms for ferrite-core amplifier with no input signal. Note: in all cases $I_d=50$ ma per cm (centimeter)

A—(At 100 kc) $V_d=50$ millivolts per cm, $V_o=10$ millivolts per cm
B—(At 500 kc) $V_d=100$ millivolts per cm, $V_o=10$ millivolts per cm
C—(At 1 mc) $V_d=200$ millivolts per cm, $V_o=20$ millivolts per cm
D—(At 3 mc) $V_d=500$ millivolts per cm, $V_o=20$ millivolts per cm

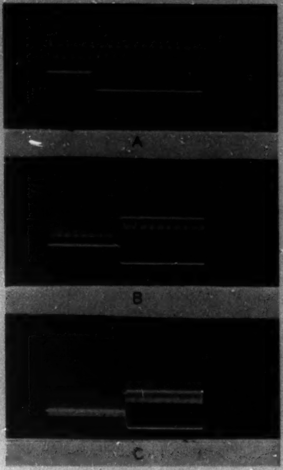


Fig. 10. Waveforms of output voltage carrier wave envelope for input current steps of increasing amplitude. Waveforms for ferrite-core amplifier. Vertical scale = 25 millivolts per cm, horizontal scale = 2 μsec per cm, excitation frequency, 1 mc

A—$I_s=1$ ma
B—$I_s=5$ ma
C—$I_s=10$ ma

The waveforms of Fig. 9 show that the presetting polarity of I_d, the excursions of I_d below the base lines in the photograph, must be increased as the frequency is raised. This is necessary because presetting must then be accomplished in a shorter time. For this reason, R_s must be decreased as the frequency is increased.

Waveforms of the output voltage carrier wave envelopes obtained when step pulses of input current are applied are shown in Fig. 10, again for the ferrite-core amplifier. The carrier increases when the signal is applied, more for one output polarity than for the other because of the asymmetric excitation. Waveforms for

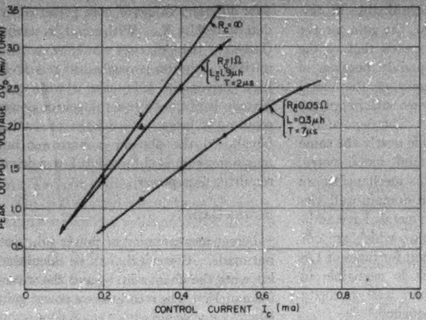


Fig. 11. Output versus input for small input signals. Curves for ferrite-core amplifier

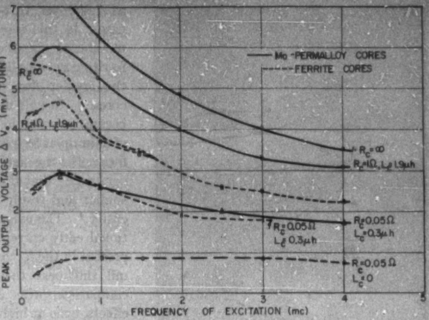


Fig. 12. Output versus excitation frequency for different input circuit R_c and L_c parameters. $I_c=1$ ma for all curves

1-, 5-, and 10-ma (milliamperes) input pulses are shown; corresponding increases in peak output are 10, 43, and 67 millivolts. These indicate some lack of output linearity for current pulses in this range of amplitude. For inputs less than 1/2 ma, however, the linearity is excellent (see Fig. 11).

The waveforms of Fig. 10 were obtained with a large value of R_c so that the L/R time constant of the input circuit was small. It is evident from the uppermost waveform of the figure that one-half cycle response is obtained in this case.

NOISE AND SIGNAL SENSITIVITY

From Fig. 10 it is apparent that signals of 1 ma can easily be detected above the no-signal carrier output level. Using care in selecting cores and in maintaining reasonably contant temperature (an oil bath was found very helpful here), sensitivities of 200 μa are possible for both the ferrite- and metal-core amplifiers. This represents an input power of less than 10^{-10} watts.

The fluctuations in output due to Barkhausen and other sources of noise were measured qualitatively. The total input equivalent noise voltage was found to be about 0.1 microvolt for both types of amplifier. This voltage was found to be nearly independent of operating conditions with the exception that it increased slightly for small values of R_c. The possibility that this increase is due to a thermal noise current circulating in the input circuit can be calculated. The rms noise current is

$$I=\sqrt{\frac{4kT\Delta f}{R}}=1.28\times10^{-10}\sqrt{\frac{\Delta f}{R}}\ \text{amperes} \qquad (1)$$

at room temprature, where k=Boltzmann's constant$=1.37\times10^{-23}$ joule per degree Kelvin (K), T=absolute temperature in degrees K, Δf=frequency range of noise, and R=total input loop resistance. Using a typical value of $R=0.01$ ohm and $\Delta f=2$ mc (see next section), one calculates $I=2$ μa which corresponds to an input voltage of 4×10^{-9} volts. Such a voltage is less than that observed by a factor of more than 10.

A better explanation for the increase in noise is that it results indirectly from the decrease in gain due to the increase in output loading as R_c is decreased (see Fig. 12). Since the Barkhausen fluctuations in flux are independent of loading, the output voltage noise is not reduced by loading, with the result that the signal-to-noise ratio decreases and the equivalent input noise current and input noise voltage increase as R_c is reduced.

The observed input noise power level is about 5×10^{-12} watts. This is a factor of 10^6 greater than that which has been obtained at low frequency[3] using large cores. The reason for this is that the Barkhausen jumps in flux are a much larger percentage of the total switching flux for the small cores used here. Reduction of the Barkhausen noise voltage by a factor of 20 by improvements in materials could reduce the noise power to a limit of about 10^{-14} watts determined by thermal fluctuations, for the case $R_c=0.01$ ohm. Accompanying improvements in zero stability would be needed, however, before this would result in a useful increase in signal sensitivity.

VOLTAGE GAIN

The steady-state voltage gain is

$$G_v=\frac{\Delta V_o}{I_cR_{in}} \qquad (2)$$

where

ΔV_o=change in peak output voltage caused by steady signal current I_c

R_{in}=input resistance of amplifier = resistance of L_c plus resistance of control winding

In the measurements of gain obtained for Figs. 5 and 12 through 16, square pulses of input current=1.0 ma were used. This means that the gain for the ferrite-core amplifier is 500 for 1 millivolt of output voltage change per output turn; for the metal-core amplifier the gain is twice as much because R_{in} is only one half as great.

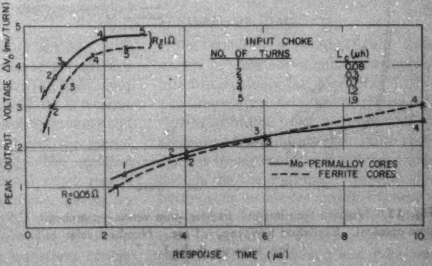


Fig. 13. Output versus response time for excitation frequency of 1 mc and input current=1-ma step. Response time for 63 percent change of output

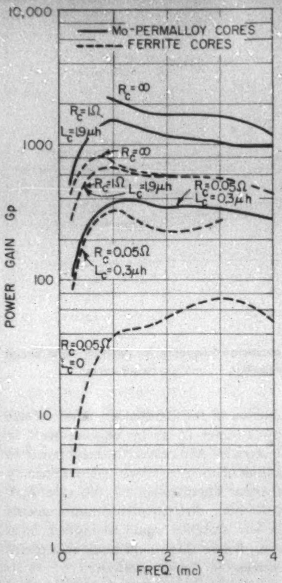


Fig. 14. Steady-state power gain G_p versus excitation frequency. G_p calculated from G_v measurements; values approximate only

For all measurements $R_L=30$ ohms was used. The core saturation excursion of I_d was also maintained constant at a peak value of 150 ma, corresponding to a magnetomotive force of 300 ma-turns.

In Fig. 12 are shown curves of ΔV_o per turn as a function of excitation frequency for different values of R_c and L_c. The greatest gain is obtained for largest R_c; this is expected because the loading effect of the input circuit upon the output is then negligible. Voltage gain as great as 7,000 per turn is obtainable in this case. In this case also, the gain for the metal-core amplifier exceeds that of the ferrite-core amplifier mainly because of the greater flux capacity of the metal cores. The gain decreases with frequency because of core losses. The dependence of gain upon frequency is nearly the same for both the ferrite and metal cores, confirming that their loss mechanisms are similar and result from domain-wall viscosity.[9] The magnetic metals have additional eddy current losses which are not present in the ferrites, but for tapes of 1/8 mil thickness this loss is negligible in comparison with domain wall viscosity effects, even at high frequency.[9]

With values of R_c less than about one ohm the gain is reduced because of output loading. The gain curves now also decline at lower frequency. This might be expected because the isolation afforded by L_c between R_c and the output decreases with frequency. Gains as great as 3,000 per turn can be obtained with values of R_c as small as 0.05 ohm and with values of L_c sufficiently small that response in less than 10 μsec can be obtained (see Fig. 13).

In Fig. 13 the dependence of V_o upon the parameters L_c and R_c is presented in a different way. In this figure ΔV_o is plotted as a function of the response time obtained using $R_c=0.05$ or 1 ohm in combination with various values of L_c. The numbers in the figure refer to the number of turns on the choke. Fig. 13 shows rather dramatically that the price paid in obtaining high gain with low source impedance is longer response time. Apparent discrepancies between the data used for Figs. 12 and 13 can be noted; these resulted because different amplifiers were used to obtain the data for each figure.

Curves of voltage gain as a function of excitation voltage are shown in Fig. 5 for different values of core presetting as determined by R_s. With small presetting (large R_s) the gain rises slowly with drive until a maximum is reached at relatively high drive; in this case a wide gain plateau is obtained but the source power required is large. With large presetting (small R_s) the plateau is narrower but maximum gain is obtained at lower drive, requiring less power.

POWER GAIN

Direct measurement of power gain was not made. Power gain can be calculated knowing the change in V_o and the waveform of V_0. The steady-state power gain is

$$G_p=\frac{1}{I_c^2R_{in}T_1}\int_{\text{cycle}}\frac{(v_o'^2-v_o^2)}{R_L}dt \tag{3}$$

where

v_o' = instantaneous output voltage with constant signal I_c
v_o = instantaneous output voltage without signal
T_1 = period of excitation

By expanding $v_o'^2-V_o^2$, equation 3 can be split into two terms:

$$G_p=\frac{1}{I_c^2R_{in}R_LT_1}\left[\int\Delta v_o^2dt+2\int v_o\Delta v_o dt\right] \tag{4}$$

where Δv_o = instantaneous change in output voltage $=v_o'-v_o$. Since v_o appears in the second part of this equation, G_p depends upon the degree of amplifier balance. With small input signals the observation was made that the waveform of v_o with and without the signal is of the same form. This means that the second term in equation 4 is positive and that unbalance increases the power gain. A lower limit for G_p can therefore be calculated by assuming perfect balance ($v_o=0$) and by numerically calculating the first term in equation 4 using the output volt-

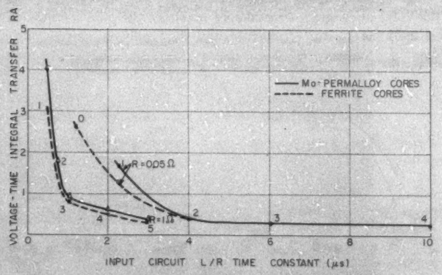


Fig. 15. Voltage time integral transfer ratio versus input circuit L/R time constant. Excitation frequency = 1 mc. Numbers refer to winding turns on choke L_c

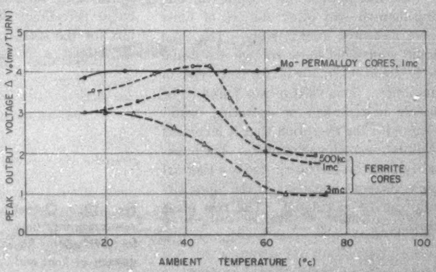


Fig. 16. Amplifier gain versus ambient temperature; $I_c=1$ ma, $R_c=1$ ohm, $L_c=0.3$ microhenry and driving current $I_d=150$ ma peak

Vol
Pt.
195

age waveshapes of Fig. 9 normalized to have peak values corresponding to ΔV_o. This was done for the output data plotted in Fig. 12, and the results are shown in Fig. 14. It should be emphasized that the values of G_p calculated in this way are only approximate because the same v_o waveshape was assumed for all values of R_L and L_c. The validity of this assumption was not critically checked.

The power gain calculations shown in Fig. 14 are for one value of load resistance only, $R_L = 30$ ohms; this value was chosen because it decreased the open circuit output voltage only moderately. Quantitative measurements to find a power maximum with R_L as a variable were not made. For this reason, and also because of the simplifying assumption of perfect balance, the values of G_p given in Fig. 14 should not be considered the maximum obtainable. The curves of G_p indicate that the power gain decreases very rapidly as R_c is reduced and that higher frequency operation is required to obtain maximum gain for small R_c. With R_c sufficiently large that output loading is negligible, a power gain maximum of greater than 2,000 is obtained at a frequency of about 1 mc.

RESPONSE

The waveforms of Fig. 10 show that one-half cycle response is obtained when R_c is large. However, for R_c a few ohms or less the L/R time constant of the input circuit limits the rate at which I_c can build up or decay to such an extent that the response time becomes greater than one-half period of the excitation. The L/R time constant then becomes the limiting factor in determining the response. For a given R_c, the fastest response results when $L_c = o$, but the loading effect of the input circuit then causes a reduction in gain as shown by the curves of Figs. 12 and 13. In practice, for each value of R_c a compromise must be made between gain and response time in choosing a value of L_c.

The inductance of the input circuit is composed of L_c plus any signal source and amplifier control winding inductance. Care was taken to maintain the signal source inductance associated with R_c very small; in this case $L = L_c + L_{in}$ only, where L_{in} is the inductance of the control winding. L_{in} was measured by observing the response with $L_c = o$ and with R_c sufficiently small that the response was much longer than one-half cycle of the excitation. Data were obtained for the ferrite-core amplifier only. With $R_c =$ 0.05 ohm, the response time obtained was 1.2 µsec measured at an excitation frequency of 3 mc for which the one-half cycle time is only 0.17 µsec. From this response time, a value of $L_{in} = 0.06$ microhenry was calculated. This value of L_{in} is negligible compared with L_c when the latter is sufficiently large to provide effective isolation.

VOLTAGE-TIME INPUT REQUIREMENTS AND VOLTAGE-TIME TRANSFER RATIO

Input voltage, current, and power comparisons have thus far been made assuming steady-state conditions. During transient conditions when I_c is changing, the instantaneous input voltage e_{in} is

$$e_{in} = L\frac{di_c}{dt} + i_c R_{in} \tag{5}$$

where i_c = instantaneous control current. If the input voltage applied is a step equal to the steady-state value $I_c R_{in}$, i_c increases according to differential equation 5 exponentially from zero to the steady-state value I_c with the time constant $L/R_{in} = (L_c + L_{in})/R_{in}$. For the amplifiers tested, this time ranged between 30 and 240 µsec. In practice, it is usually desirable to supply more than the steady-state input voltage during the transient period to support a greater inductive drop and therefore to increase di_c/dt and reduce the response time. This means that the instantaneous voltage gain will be less than the steady-state gain. An exact calculation of the instantaneous gain requires a specification of the waveform of the input voltage applied and in general is complicated.

Another way to consider the transient operation is in terms of the volt-second transfer ratio.[10] This is defined as the change in the half-cycle output voltage-per-turn time integral divided by the change in the input voltage-per-turn time integral required to produce the output change. It is, therefore, a measure of the output flux change produced by an effective input flux change. The input integral is obtained by integrating equation 5:

$$\int_{T_2} e_{in}dt = LI_c + R_{in}\int_{T_2} i_c dt \tag{6}$$

where T_2 = duration of the transient response.

Since fast response ($T_2 << L/R_{in}$) is usually desired in practice and is obtained by considerably overdriving the input during the transient period, the first terms on the right-hand sides of equations 5 and 6 are then predominant and the second terms can be neglected. The modified input volt-second integral which results is thus that for an amplifier having an ideal control winding of zero resistance. This integral, not involving i_c, is independent of input waveform parameters and depends only upon L and the steady-state value of control current. The transfer ratio calculated using the simplified input integral then is a simple constant which can be used to calculate input volt-second requirements given specified output conditions. Curves of voltage-time integral transfer ratio calculated using this simplification are given in Fig. 15, utilizing the data for Fig. 13 and the waveforms of Fig. 9. The assumption that the output waveform with signal is of the same form as the waveform without signal was again made in calculating the output voltage-time integrals.

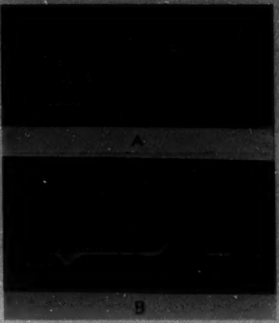


Fig. 17. Partial switching output voltage waveforms for 10 ferrite and 10 metal tape cores

A—(Ferrite cores) vertical scale = 5 millivolts per cm, horizontal scale = 0.2 µsec per cm

B—(Mo-Permalloy cores) vertical scale = 16 millivolts per cm, horizontal scale = 0.5 µsec per cm

The voltage-time integral transfer ratio curves of Fig. 15 show that transfer ratios greater than 1.0 are obtained if L_c is sufficiently small. That such flux gains are possible at high frequency has also recently been shown by Crane[11] in experiments with coupled magnetic logic elements.

TEMPERATURE STABILITY

The stability of voltage gain with respect to variation in ambient temperature was measured (see Fig. 16). To obtain temperature data, the amplifiers, including mounting bases, were operated in a constant-temperature oven; core cooling was by air convection and heat conduction via the windings. The gain stability of the metal-core amplifier is nearly constant over the range indicated, which is to be expected. The gain stability of the ferrite-core amplifier is also quite con-

Table I. Zero Balance and Signal Sensitivity for a Number of Amplifiers Built Using Cores Selected at Random

		$R_s = \alpha$, $L_e = o$ Balance $(p\text{-}p)\Delta V_o$		$R_e = 1$ ohm, $L_e = o$ Balance $(p\text{-}p)\Delta V_o$	
Ferrite Core Amplifier No.	1	13.5	2	11	1.7
	2	7	3.2	6	2.2
	3	7.5	2	6	1.5
	4	6	3.5	5	3
	5	5	3.7	3.5	2.3
Molybdenum Permalloy Core Amplifier No.	1	24	8.7	17	7
	2	8	11	7	8
	3	5.3	11	5.3	7
	4	19	8	19	7
	5	37	7	28	5.3

Peak Driving Current = 150 ma
$R_L = 30$ ohms, R_s adjusted for maximum sensitivity
Output in millivolts per turn

Frequency = 1 mc
$I_e = 1$ ma

stant over a narrower temperature range (20-50 C) for excitation frequencies of 1 mc or less.

The amplifier zero balance is also dependent upon temperature and excitation frequency. Quantitative measurements of zero drift were not made, but it appears to be satisfactory over the same range of temperature and frequency for which gain stability is obtained.

The gain and zero stability for the ferrite amplifier, especially at higher frequencies of excitation, is considerably improved by immersing the amplifier in cooling oil. Quantitative data concerning this were not obtained.

UNIFORMITY

Amplifiers were constructed using pairs of cores picked at random and compared for gain and zero balance. The uniformity of balance was found to be considerably better for the ferrite-core than for the metal-core amplifiers. To investigate this further, pulse switching data were obtained for a number of cores of each type picked at random; superimposed switching output waveforms are shown in Fig. 17. The cores were driven to saturation by one polarity of drive and preset approximately 50% by the other polarity which was of constant amplitude. The uniformity of flux switched under these conditions is better for the ferrite than for the metal cores.

In Table I are given balance and sensitivity data for five amplifiers of each type constructed using randomly selected cores and operated at room temperature with air-convection cooling. Results are also given for two values of R_e. It is probable that the gain and zero balances vary too much for many applications; for these better uniformity can be obtained by core matching.

The molybdenum permalloy cores used here were chosen from a different lot than those used in obtaining the data reported earlier in the paper and exhibited a lower coercive force and a poorer squareness. The amplifier sensitivities with these cores were greater by a factor of more than two compared to the sensitivities for the metal cores used earlier.

Magnetic Memory Sensing Application

The very low input impedance of this type of amplifier and its great voltage sensitivity lead one to suggest its use as an input mixing sensing amplifier for random-access magnetic memories. Conventional magnetic-core memories have typical access times of about 10 μsec and require a sensing amplifier response time of 1 or 2 μsec. This type of magnetic amplifier will satisfy such a response requirement if its input L/R time constant is small enough. A possible system is outlined in Fig. 18. Sensing of a single digit plane core array of a memory is shown as representative. Information storage and interrogation is done conventionally; e.g., by X-Y current coincidence, as shown in the figure, or by a linear mode of selection. However, rather than sensing conventionally by using a single transistor- or vacuum-tube amplifier coupled to a sensing winding which links every core of the array, in this system the sensing winding is split into shorter windings each sensing only a few cores and the outputs of these are mixed together in groups (via the resistances R indicated in the figure) and each group is fed into a magnetic amplifier. Only a limited number of sensing winding outputs can be mixed per magnetic amplifier without slowing down the response of the amplifier too greatly due to input shunting. Further mixing is done by series connecting the outputs of the magnetic amplifiers to a common output transformer. In this latter mixing operation, only one of the amplifiers is excited at a time (excitation of A or B or C in Fig. 18) to avoid the cumulative effects of amplifier zero unbalances.

The noise problem will be considered next. During the memory-reading operation many unselected memory cores generate small disturb signals. In each sensing winding these disturbs can be made to cancel by favorable choice of a sensing polarity configuration, but statistical fluctuations in the disturbs cause this cancellation to be imperfect and small net disturb signals usually result which may be additive in the amplifier input mixing circuits. To maintain a satisfactory signal-to-noise ratio in each amplifier, it is therefore necessary, for this reason (as well as for the response time reason previously mentioned), to limit the number of sensing winding outputs which are mixed into a single amplifier. A satisfactory over-all signal-to-noise ratio is then obtained by exciting only the one magnetic amplifier of the array which is being signalled by the wanted storage information. All unexcited amplifiers remain inactive and do not contribute noise in the output circuit loop. To do this,

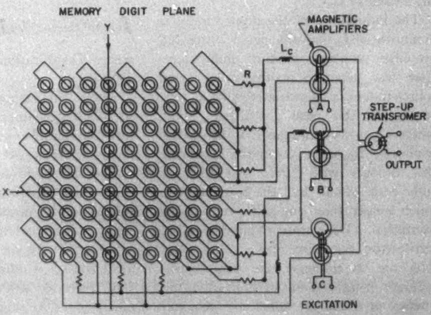


Fig. 18. Use of magnetic amplifiers in sensing random-access memory

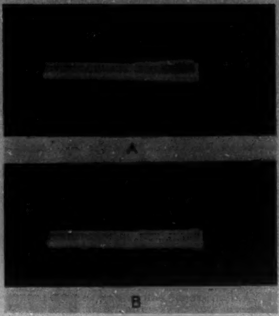


Fig. 19. Magnetic sensing amplifier output waveforms

A—"1" information: vertical scale = 10 millivolts per cm
B—"0" information: vertical scale = 10 millivolts per cm

additional logical circuitry associated with that of memory addressing for locating the correct amplifier to excite is necessary, and the excitation circuits must each transmit sufficient power to drive the amplifier cores adequately. Magnetic logic devices appear to be promising for this purpose, especially those using transfluxors to gate the carrier.[12]

The necessary system with magnetic sensing of using split sensing windings has another advantage. A sensing winding which links many cores acts as a delay line. An information signal generated at the far end of such a line may be seriously delayed in reaching the amplifier and, because the delay depends upon the interrogated location, time strobing of the output may be difficult. By splitting the sensing winding into shorter sections, the signal delay in reaching the amplifier is greatly reduced and the delay becomes predominantly the constant delay due to the response of the magnetic amplifier itself. Transmission in the amplifier output loop could introduce another variable delay, but in practice this is negligible because relatively few magnetic amplifiers are coupled and single-turn output loops can be used to reduce the delay per amplifier. Single-turn loops result in low voltage gain, but this can be increased by the use of a step-up transformer as shown in Fig. 18.

The design steps to determine the number of cores which can be sensed per winding and the number of windings which can be mixed per amplifier are outlined in the following.

The resistance R must first be chosen to obtain sufficient signal current to be detected by the amplifier from the "1" state voltage generated by the memory core during interrogation. However, this current must not exceed a value which could alter the information state of other cores. The number of cores which can be linked by a single sensing winding is next found, considering the contribution to R due to the winding resistance and the signal voltage drop due to inductive effects in the cores. External resistance can be eliminated if the winding resistance itself is made equal to R. The number of sensing windings which can be mixed are finally determined. Aside from the disturb noise consideration, the main consideration here, as already mentioned, is the response time. The shunt resistance of the paralleled sensing winding circuits must be great enough to yield a sufficiently small response time for the L_c value used. Signal-current shunting by the inactive windings is not usually important because the amplifier input resistance is in general negligible compared to the shunting resistance. If long response times are permissible in the design, however, signal shunting should be also be considered.

For the amplifiers to have good gain, some input choke L_c is required. This is usually lumped, but in some cases the self-inductances of the sensing windings themselves are sufficiently large that their paralleled inductance is enough to reduce the output loading to a tolerable value without the need for additional series inductance in the control circuit.

During excitation, carrier-frequency voltages induced on the control windings of the amplifiers cause carrier-frequency currents to flow in the sensing windings. These currents, if sufficiently large, could alter the information states of the cores. However, in practice, because of the choking action of L_c and the paralleling of sensing windings, the individual currents flowing in each winding are negligible.

In Fig. 19 are shown typical waveforms of the "1" and "0" information output signals obtained from a ferrite core sensing amplifier. The input signal was furnished by a small slow-switching memory core for purposes of experimental convenience. The excitation frequency was 500 kc, the horizontal sweep 10 μsec per centimeter, and $R_c = 0.3$ ohm. The peak value of control current was 1 ma.

Another advantage of magnetic amplifier sensing with respect to conventional methods is the capacity of this type of amplifier to recover quickly after a large input signal has been applied. Such inputs can occur when inhibiting current pulses are applied to the memory core array during the information writing portion of the memory cycle. With transistor or vacuum-tube low-level amplifiers having several capacitance coupled stages the recovery time may be prohibitively long. To obtain sufficiently fast recovery d-c coupling must usually be resorted to and this introduces zero-drift difficulties. With magnetic amplifiers on the other hand, only a single stage is needed and the output response is sufficiently nonlinear that the amplifier is self-limiting. This action is illustrated with the aid of the waveform photograph of Fig. 20. This shows the output carrier wave envelope obtained for a 1-ma input pulse applied soon after a saturating input pulse of 100 ma. The input L/R time constant of the amplifier was 4 μsec. This time constant, although too large for a practical system, was chosen for demonstration purposes. Small-signal response was nearly complete in two time constants, or about 8 μsec. From the figure it can be seen that the recovery of the amplifier was nearly complete about 20 μsec after the termination of the saturating pulse (both the output envelope and the current pulse waveforms were off scale on the oscilloscope during the saturating pulse interval). This recovery time is only slightly greater than twice the small signal response time. Recovery in less than 20 μsec to make the device practical could of course be obtained by reducing the L/R input time constant.

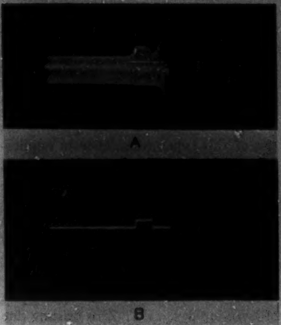


Fig. 20. Recovery of magnetic amplifier after application of 100-ma input pulse. Time scale = 20 μsec per cm

A—Output voltage = 10 millivolts per cm
B—Input current pulses = 2.5 ma per cm

Conclusions

High-frequency magnetic amplifiers of magnetic modulator design, asymmetrically excited and constructed of small ferrite or metal tape cores, show promise for stable amplification of signals as small as 10^{-19} watts with response times of 10 μsec or less.

A voltage sensitivity of less than 1 μv can be obtained which permits the use of this type of amplifier in applications, such as computer memory sensing, in which considerable signal voltage division occurs because of mixing. High voltage gain and a moderate power gain can be obtained even with only a few turns on the output winding. Carrier gating of the amplifier is feasible. Recovery is rapid after the application of large input signals. Zero drifting due to heating of the magnetic cores during excitation and the necessity for core matching to obtain best zero balance are difficulties for which complete solutions can undoubtedly be found.

References

1. THE STUDY OF A MAGNETIC INVERTER FOR AMPLIFICATION OF LOW-INPUT POWER D.C. SIGNALS, **E. H. Frost-Smith.** *Proceedings*, Institution of Electrical Engineers London, England, vol. 100, pt. 2, 1953, pp. 362–75.

2. AN EVEN-HARMONIC MAGNETIC AMPLIFIER, **P. D. Atkinson, A. V. Hemingway.** *Electronic Engineering*, London, England, vol. 26, no. 11, Nov. 1954, pp. 482–5.

3. THE FUNDAMENTAL LIMITATIONS OF THE SECOND-HARMONIC TYPE OF MAGNETIC MODULATOR AS APPLIED TO THE APPLICATION OF SMALL D.C. SIGNALS, **F. C. Williams, S. W. Noble.** *Proceedings*, Institution of Electrical Engineers, vol. 97, pt. 2, 1950, pp. 445–59.

4. THE TRANSACTOR: A SELF-SATURATED TRANSFORMER-PART I, **A. B. Rosenstein.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, May 1958, pp. 129–41.

5. AN IMPROVED MAGNETIC SERVO AMPLIFIER, **C. W. Lufcy, A. E. Schmid, P. W. Barnhart.** *Ibid.*, vol. 71, Sept. 1952, pp. 281–89.

6. DESIGN CRITERION FOR LOW-LEVEL SECOND-HARMONIC MAGNETIC MODULATORS, **E. J. Kletsky.** *Ibid.*, vol. 77, 1958 (Jan. 1959 section), pp. 1013–19.

7. A SIMPLE MAGNETIC MODULATOR FOR CONVERSION OF MILLIVOLT D-C SIGNALS, **G. Wennerberg.** *Electrical Engineering*, vol. 70, Feb. 1951, pp. 144–47.

8. MAGNETIZATION IN TAPE-WOUND CORES, **R. C. Barker.** *Proceedings*, AIEE Special Technical Conference on Non-Linear Magnetics and Magnetic Amplifiers, 1958, pp. 1–46.

9. MAGNETIC MATERIALS FOR DIGITIAL COMPUTER COMPONENTS, **N. Menyuk, J. B. Goodenough.** *Journal of Applied Physics*, New York, N. Y., vol. 26, 1955 pp. 8–18.

10. VOLT-SECOND TRANSFER EFFICIENCY IN FAST-RESPONSE MAGNETIC AMPLIFIERS, **T. J. Pula.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, 1958 (Jan. 1959 section), pp. 861–67.

11. A HIGH SPEED LOGIC SYSTEM USING MAGNETIC ELEMENTS AND CONNECTING WIRE ONLY, **H. D. Crane.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 47, 1959, pp. 63-73.

12. TRANSFLUXOR CONTROLLED ELECTROLUMINESCENT DISPLAY PANELS, **J. A. Rajchman, G. R. Briggs, A. W. Lo.** *Ibid.*, vol. 46, 1958, pp. 1808–24

A Mathematical Model for a Magnetic-Amplifier Reactor Core

H. H. WOODSON
ASSOCIATE MEMBER AIEE

Vol. 7
Pt. 1
1959

IN THE ANALYSIS and design of self-saturating magnetic amplifiers, the flux-resetting characteristics of the saturable reactors must be known either analytically or graphically. In the past, the reactor flux-resetting characteristic has been represented by a single, unique relation between flux and resetting magnetomotive force.[1–3] Such a restrictive reactor representation has limited usefulness because changes in operating conditions such as control voltage waveform, relative control circuit impedance level, and supply frequency cause considerable change in the flux-resetting characteristic.

More recently, several studies have been aimed at improving the representation for a saturable reactor to include the dynamic properties. Several of these studies[4–8] have dealt with the experimental behavior of saturable reactors over a wide-range of operating conditions. The principal shortcoming of these studies is that they are purely experimental; consequently, without a theoretical model to use for extrapolation, experimental information must be obtained for every conceivable operating condition. Friedlander has been concerned with the prediction of flux-resetting characteristics by using ferromagnetic theory and making some assumptions about the statistical nature of a polycrystalline sample.[9] The types of mathematics encountered in this treatment restrict the analyses to simple, constant-current resetting characteristics, thus, the results are applicable to a restricted class of magnetic-amplifier problems. Still other workers have determined or assumed empirical expressions describing flux resetting and have used them in magnetic-amplifier analyses.[10] Such an approach suffers from the lack of a model of the physical processes that would reduce the amount of experimentation necessary for the description of a given problem.

This paper will describe the derivation of a mathematical model for a saturable reactor core. The starting point is domain theory of ferromagnetism and the object is equations which are simple enough to have engineering applicability. In essence, the approach is the following: 1. flux reversal is assumed to occur by the growth of a large number of domains; 2. the walls of each domain are assumed to move against linear damping forces; 3. the domains are assumed to be uniformly distributed throughout the material; 4. all domains are assumed identical except for their starting fields; 5. the starting fields are assumed to have some distribution function. This approach has the elements of the one cited previously by Friedlander,[9] who assumed that all domains are identical in all respects but that there is a distribution of domain positions throughout the sample. In addition he assumed all nucleation of domains to occur at material surfaces. Friedlander's treatment and this paper represent only two of a variety of ways in which the behavior of a single domain can be averaged over some statistical property to obtain a representation for a polycrystalline sample.

In any treatment of this sort the assumptions are of necessity speculative because of the present state of knowledge in the field. The use of these assumptions to obtain plausible explanations for observed phenomena not only serves present day engineering needs but also indicates areas where fundamental research may yield useful results in either verifying or proving invalid the various assumptions.

Paper **59-1058**, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-IRE Nonlinear Magnetics and Magnetic Amplifiers Conference, Washington, D. C., September 23–25, 1959. Manuscript submitted November 12, 1958; made available for printing July 17, 1959.

H. H. WOODSON is with the Massachusetts Institute of Technology, Cambridge, Mass.

This paper is taken, in part, from a thesis submitted to the Electrical Engineering Department at Massachusetts Institute of Technology on May 20, 1956, in partial fulfillment of the requirements for the Sc.D. degree. The work was conducted with the Energy Conversion Group of the Servomechanisms Laboratory at Massachusetts Institute of Technology and was supported, in part, by Wright Air Development Center under Contract *AF33(616)3242.*

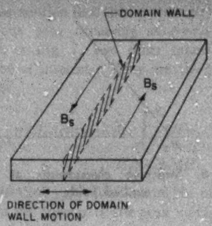


Fig. 1. Illustration of domain wall movement

Theoretical Basis for Model

The starting point for the derivation of the mathematical model for a polycrystalline ferromagnetic material is domain theory of ferromagnetism.[11–13] In this theory a material consists of domains in which the material is magnetically saturated. The directions of the magnetization in the domains are the directions of easy magnetization of the material. The domains are separated by domain walls through which the direction of magnetization changes abruptly. When a specimen is completely saturated in one direction, only one domain covering the entire specimen exists; when a reverse field is applied to reverse the magnetization, small domains of reverse magnetization form and their walls move irreversibly perpendicular to the direction of magnetization to effect a flux reversal. This is illustrated in Fig. 1, where B_s represents the saturation flux density.

From this discussion it is apparent that there are three characteristics of a domain that must be described: 1. The field necessary to form or nucleate domains of reverse magnetization in a sample that has been previously strongly saturated. 2. The minimum field (starting field) necessary to make the walls of the domain move through the material to change the net flux. 3. The mechanisms which impede the motion of a domain wall through the material.

The process of nucleation is not well understood although there have been theoretical[14,15] and experimental[13,16] studies to determine some of the characteristics of the process. The assumptions to be made here are given in the next section.

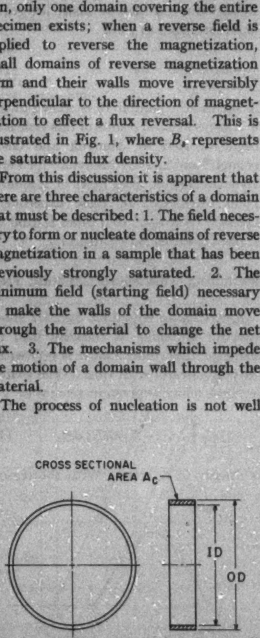


Fig. 2. Geometry of specimen for analysis

It is well accepted that the motion of an element of an existing domain wall can be described by the equations[16–18]

$$f(\delta)\frac{d\delta}{dt}=K_1(H-H_s),\ \text{for}\ H>H_s \tag{1}$$

where

δ = distance of domain wall from nucleating site
$f(\delta)$ = position dependent damping factor
K_1 = material constant
H = applied magnetic field
H_s = starting field

Equation 1 is simply a first-order differential equation expressing the fact that the force that drives the wall (right-hand side) is all absorbed by a dissipative damping mechanism whose factor $f(\delta)$ depends on position.

In the light of present understanding, equation 1 should include an inertia term; however, it has been shown experimentally that inertia terms become appreciable only at much higher frequencies than are of interest here.[19]

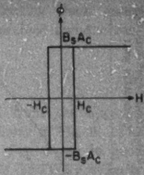
In general the several factors in equation 1 are impossible to describe in detail for every element of every domain wall in a polycrystalline material. For instance the damping factor not only depends on spin relaxation damping which is thought to be an inherent property of a material[15,16] but also on macroscopic eddy currents which in turn depend on the detailed configuration of all domain walls relative to material surfaces. The reduction in local magnetic field caused by macroscopic eddy currents is included in the $f(\delta)(d\delta/dt)$ term. The field H_s necessary to start the domain wall moving depends mostly on the detailed structure of a material because domain walls tend to become attached to imperfections. Thus, for generality, the starting field should be a function of wall position. Since little is known about the nature of these quantities, plausible assumptions will be made in the next section.

When an element of a domain wall moves it changes the flux density from the saturation value B_s in one direction, to B_s in the other direction. Thus the rate of change of flux due to motion of the element is

$$\frac{d\phi_1}{dt}=2B_s\frac{dA_1}{dt} \tag{2}$$

where

Fig. 3. Hysteresis loop assumed for material



$\frac{d\phi_1}{dt}$ = contribution to total rate of change of flux made by domain wall element
B_s = saturation flux density
$\frac{dA_1}{dt}$ = rate at which area perpendicular to domain wall is swept out by element of wall

Equations 1 and 2 form the theoretical basis for the mathematical model to be derived. The type of information necessary for applying these expressions to the analysis of a polycrystalline sample is not available; consequently, assumptions must be made to complete the model.

Assumptions

1. The specimen to be analyzed is the toroidal geometry shown in Fig. 2. The inside-diameter–outside-diameter ratio is assumed very near unity so that the applied field is uniform throughout the material. The easy direction of magnetization is around the circumference; thus, all magnetization will be in this direction and domain walls will move radially or axially.
2. The nucleating field for each domain is less than its starting field.
3. The d-c hysteresis loop for the specimen is rectangular as shown in Fig. 3. That is, if a field infinitesimally greater than H_c is applied when the material is saturated negatively, the flux will reverse completely. When the specimen is completely magnetized in either direction, no further flux change can occur.
4. The rectangularity of assumption 3 states that at least one domain has a starting field H_c (coercive field) which remains constant throughout the material, and that H_c is the minimum starting field for the material. It is furthermore assumed that all domains have starting fields $H_s > H_c$ which are constant throughout the domain wall motion.
5. All flux changes occur irreversibly.
6. Before any flux change, the material has been saturated strongly to establish uniform initial conditions for a subsequent flux change. This assumption agrees well with conditions found in most self-saturating magnetic amplifiers.

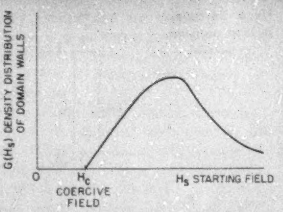


Fig. 4. Distribution of starting fields

7. All domains are assumed identical and uniformly distributed over the cross section of the specimen except that the starting field H_s varies among domains according to some distribution function which is independent of applied field.
8. Each domain is described by equations of the form of equations 1 and 2.
9. The domain wall damping factor $f(\delta)$ is independent of wall position, $f(\delta)=1$ in equation 1.
10. The time rate of change of applied field is small so that no large, abrupt changes in the amount of domain wall occur during a flux change. This assumption disagrees with the type of operation found in full-wave self-saturating magnetic amplifiers.

The Mathematical Model

To complete the mathematical model with the use of equations 1 and 2 and the above assumptions, some averaging must be done. In assumption 7 it is stated that the starting field for the domain walls varies over some distribution. Assume that this distribution function defined by $G(H_s)$ is some general curve shown in Fig. 4. The function $G(H_s)$ expresses how many elements of domain wall have starting fields between H_s and H_s+dH_s.

Now, multiply both sides of equation 1 by $G(H_s)$ and integrate over H_s to obtain

$$\frac{d}{dt}\int_{H_c}^{H}\delta G(H_s)dH_s=\int_{H_c}^{H}K_1(H-H_s)G(H_s)dH_s \qquad (3)$$

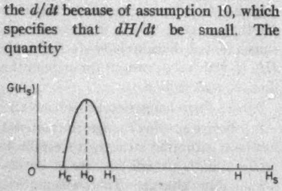


Fig. 5. One type of distribution of starting fields

The integral on the left is taken inside the d/dt because of assumption 10, which specifies that dH/dt be small. The quantity

$$\int_{H_c}^{H}\delta G(H_s)dH_s$$

represents the total distance over which all domain wall elements move. For a complete flux reversal this quantity will proceed from zero to some constant value no matter how the reversal comes about, provided the assumptions are satisfied. Thus an average normalized domain dimension r can be defined as

$$r=K_2\int_{H_c}^{H}\delta G(H_s)dH_s \qquad (4)$$

with the constant K_2 defined to make $r=$ unity, when the flux is completely reversed. Substitution of equation 4 into equation 3 yields

$$\frac{dr}{dt}=K_1K_2\int_{H_c}^{H}(H-H_s)G(H_s)dH_s \qquad (5)$$

This is one equation for the general mathematical model.

When equation 2 is integrated over all domain wall elements that are moving, there results

$$\frac{d\phi}{dt}=2B_s\frac{dA_d}{dt} \qquad (6)$$

where
ϕ=total flux
A_d=total area swept out by moving domain walls

At the start of a flux reversal ($r=0$), $A_d=0$; and at the completion of a flux reversal ($r=1$), $A_d=A_c$.

An additional assumption is that as the area A_d changes it is a function of only the average normalized domain dimension r; consequently, equation 6 can be rewritten as

$$\frac{d\phi}{dt}=2B_sA_cg(r)\frac{dr}{dt} \qquad (7)$$

where $g(r)$ is a function of r which has the property

$$\int_0^1 g(r)dr=1 \qquad (8)$$

The function $g(r)$ can be interpreted as a "shape function." That is, when the applied field H is constant in equation 5, r varies linearly with time; and, from equation 7, $g(r)$ represents the shape of the voltage pulse for constant field switching.

Equations 5 and 7 are the mathematical model for the polycrystalline material. Before this model can be used, the functions $G(H_s)$ and $g(r)$ must be defined. These functions will now be considered.

Switching with Large Constant Field

Consider the case of a material in which the range of starting fields is limited and the distribution function is as shown in Fig. 5. When the applied field is sufficiently large, the integration in equation 5 can be simplified by assuming $G(H_s)$ to be an impulse at $H_s=H_0$, the value of field that divides the area under the curve in half, and having an area equal to the area under $G(H_s)$. Such an approximation yields

$$\frac{dr}{dt}=K_1K_2\int_{H_c}^{H}(H-H_s)G(H_s)dH_s=K_1K_2N_0(H-H_0) \qquad (9)$$

where N_0 is the area under the $G(H_s)$ curve.

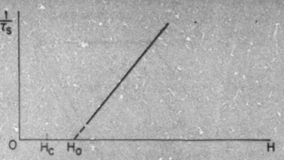


Fig. 6. Switching characteristic at high applied fields

Now consider the time τ_s required to switch the flux in this reactor when the applied field H is constant and $H>>H_1$. Equation 9 can then be easily integrated and rearranged to yield

$$\frac{1}{\tau_s}=K_1K_2N_0(H-H_0) \qquad (10)$$

A sketch of equation 10 is shown in Fig. 6. Experimental results of this type have been reported by Menyuk and Goodenough[18] and later by Menyuk[20] for Permalloy cores made from tapes of 1/8-mil to 1-mil thickness. In these experiments the field H_0 was 1.5 to 2 times H_c and the applied field ranged up to 10 H_c.

Switching with Small Constant Field

Reference to equations 5 and 7 shows that for switching at constant field the shape of the voltage pulse is $g(r)$ with r varying linearly with time. The switching time which is given in general by

$$\frac{1}{\tau_s}=K_1K_2\int_{H_c}^{H}(H-H_s)G(H_s)dH_s \qquad (11)$$

is usually of interest. With the reason-

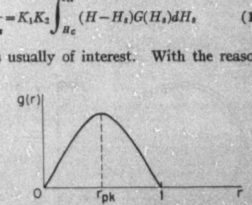


Fig. 7. Shape of voltage pulse for constant applied field

able assumption that the function $g(\tau)$ has only one peak occurring at $\tau=\tau_{pk}$, as shown in Fig. 7, two other quantities may be of interest. They are the peak voltage

$$\left(\frac{d\phi}{dt}\right)_{pk}=2B_sA_cg(\tau_{pk})K_1K_2\int_{H_c}^{H}(H-H_s)G(H_s)dH_s \quad (12)$$

and the time τ_{pk} required from the initiation of the pulse until the peak voltage is reached:

$$\frac{1}{\tau_{pk}}=\frac{K_1K_2}{\tau_{pk}}\int_{H_c}^{H}(H-H_s)G(H_s)dH_s \quad (13)$$

When the applied field has a value in the range where $G(H_s)$ has an appreciable value, no simple linear relation such as equation 10 results. Instead, the function $G(H_s)$ must be known and equations 5 and 7 used without simplification. In the absence of other information from which to obtain $G(H_s)$, it can be inferred from measurements of constant field-switching characteristics.

Any shape of function $G(H_s)$ can be represented to any degree of accuracy over any range of fields by the power series

$$G(H_s)=\sum_{n=0} a_n(H_s-H_c)^{n+p} \quad (14)$$

where

$H_s>H_c$
a_n are constants
$-1<p<0$
n is an integer

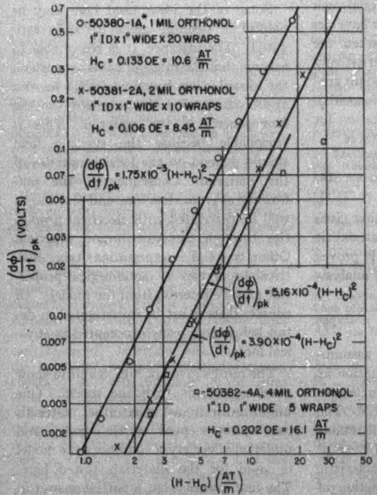


Fig. 8 (left). Plots of $(d\phi/dt)_{pk}$ versus $(H-H_c)$ for three reactors

The peak voltage (equation 12) will be used subsequently, but the process that follows can just as easily be used with τ_s and τ_{pk}.

Substitution of equation 14 into equation 12 and integration of the result yield

$$\left(\frac{d\phi}{dt}\right)_{pk}=2B_sA_cg(\tau_{pk})K_1K_2\sum_{n=0}^{\infty}\frac{a_n}{(n+p+1)(n+p+2)}(H-H_c)^{n+p+2} \quad (15)$$

Constants a_n and p could be chosen to make this expression fit very closely a curve of $(d\phi/dt)_{pk}$ versus $(H-H_c)$; however the resulting model would be difficult to apply in any analytical problems. It proves more practical to obtain an approximate fit with experimental curves in which only one term in the series is retained.

To show that this is so in at least one case, consider the experimental curves of Fig. 8 which were taken with the use of the circuit of Fig. 9. When switch S, which was a set of parallel beam power tubes (*6 L 6*), was closed, the core was strongly saturated. When switch S was opened, the reset source V_r reset the core through the resistance R_r, which was so large that the current i_r remained essentially constant during the resetting process. It is evident from Fig. 8 that these curves are represented to a good degree of accuracy by an $(H-H_c)^2$ term. Consequently, for representing these cores, $(n=p=0)$ can be specified and the constant A defined by

$$A=B_sA_cg(\tau_{pk})K_1K_2a_0 \quad (16)$$

can be determined from the approximate solid curves shown in Fig. 8. The results obtained in Fig. 8 are in good agreement with those reported previously by Huhta.[8]

It is interesting to note that with $n=p=0$ the distribution function $G(H_s)$ is a constant over the range of fields considered. Such an approximation is shown in Fig. 10.

Switching with Slowly Varying Field

To show that the mathematical model is applicable to at least one case in which the applied field varies, and is therefore applicable to some types of actual magnetic-amplifier circuits, consider the resetting of a reactor with the circuit of Fig. 9 in which the reset resistance R_r is small enough that the reset current i_r varies appreciably during the resetting process. To obtain a comparison with experimental results, assume the reactor to have a constant-current switching characteristic that varies with $(H-H_c)^2$, as shown in Fig. 8. Then the mathematical model of equations 5 and 7 reduces to

$$\frac{d\tau}{dt}=K_1K_2\frac{a_0}{2}(H-H_c)^2 \quad (17)$$

$$\frac{d\phi}{dt}=B_sA_cg(\tau)K_1K_2a_0(H-H_c)^2 \quad (18)$$

The voltage equation for the reset circuit is

$$V_r=i_rR_r+N_r\frac{d\phi}{dt} \quad (19)$$

Multiplying this expression by N_1/l_mR_r yields

$$\frac{N_rV_r}{l_mR_r}=H+\frac{N_r^2}{l_mR_r}\frac{d\phi}{dt} \quad (20)$$

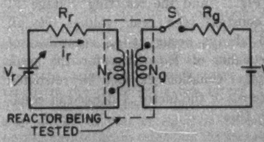


Fig. 9 (right). Circuit for checking model with small constant field and with slowly varying field

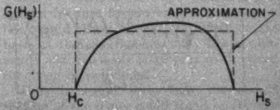


Fig. 10 (right). Approximation for function G(H) to yield switching characteristics that vary as the square of the effective field

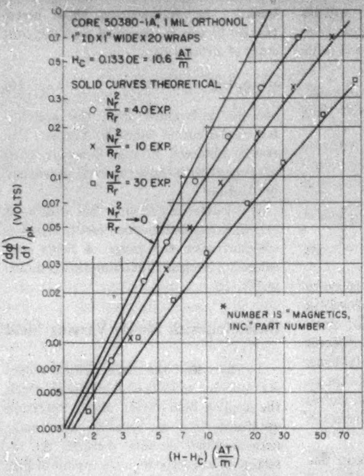


Fig. 11. Same as Fig. 8, for various values of N_r^2/R_r

where

$$H=\frac{N_r i_r}{l_m} \tag{21}$$

is the instantaneous applied field, and l_m is the mean length of the magnetic path.

Assuming that the switch S in Fig. 9 is closed at $t=0$, with $r=0$, simultaneous solution of equations 17, 18, and 20 yields

$$\frac{d\phi}{dt}=B_sA_cq(r)K_1K_2a_0\times \left[\frac{-1+\sqrt{1+4B_sA_cg(r)K_1K_2a_0\frac{N_r^2}{l_mR_r}\times \left(\frac{N_rV_r}{l_mR_r}-H_c\right)}}{2B_sA_cg(r)K_1K_2a_0\frac{N_r^2}{l_mR_r}}\right]^2 \tag{22}$$

Whenever the function $g(r)$ has only one peak occurring at r_{pk}, as shown in Fig. 7, and the resetting voltage V_r is constant, the peak voltage $(d\phi/dt)$ will always occur at r_{pk} regardless of the value of R_r. Thus for the peak voltage, equation 22 can be rewritten in terms of the factor A defined in equation 16 as

$$\left(\frac{d\phi}{dt}\right)_{pk}=A\times \left[\frac{-1+\sqrt{1+4A\frac{N_r^2}{l_mR_r}\left(\frac{N_rV_r}{l_mR_r}-H_c\right)}}{2A\frac{N_r^2}{l_mR_r}}\right]^2 \tag{23}$$

The factor A determined from the constant current switching curves of Fig. 8 for the 1-mil core is used with equation 23 to predict curves for three values of N_r^2/R_r, as shown in Fig. 11. Experimental points are shown for the same conditions, and the agreement is good.

Further Simplifications

So far nothing has been said about the detailed shape of the function $g(r)$. A high-performance magnetic material such as the 50Ni-50Fe material will have an almost rectangular **B-H** loop when the flux is driven sinusoidally. Alternatively when the flux is switched at constant current the voltage pulse is almost sinusoidal. Consequently, a good approximation for so-called rectangular loop material is

$$g(r)=\frac{\pi}{2}\sin \pi r, \text{ for } 0\leq r\leq 1 \tag{24}$$

Although such an approximation gives a good representation of the material, it makes any analysis difficult.[21] It proves more profitable in many cases of analysis to assume that $g(r)$ is a constant:

$$g(r)=1, \text{ for } 0\leq r\leq 1 \tag{25}$$

For instance, such a simplifying assumption allows the application of equation 22 directly to the analysis of a single-core self-saturating magnetic amplifier.[21] The simplification of equation 25 still retains the essential nonlinearity of the $(H-H_c)^2$ variation in the switching characteristic and therefore allows good prediction of single-core self-saturating magnetic-amplifier characteristics over a wide range of operating conditions, e.g., N_r^2/R_r frequency.[21]

A further simplification that can be made for operation at one frequency is to assume a $g(r)$ given by equation 25 and to linearize equation 17 and 18 about some mid-range value of applied field H. Such a simplification results in the representation of the saturable-reactor exciting characteristics as a constant-current source shunted by a constant resistance. This approximation which leads to extremely simple and easily interpretable magnetic-amplifier analyses is illustrated in a compansion paper by Frankenthal.[22] This linear approximation can easily be correlated with the incremental permeability defined by Roberts[4] for constant current resetting. Thus, the results of a Roberts core test can be used with this linear model to predict magnetic-amplifier performance for other than constant-current reset.

Results and Conclusions

In the derivation of the mathematical model, so many speculative assumptions have been made that some readers may doubt that there is any connection between the theory of ferromagnetism and the resulting model. With the present state of knowledge, this point can be argued but not settled. It is the hope of the author that arguments and further experiments will ensue to shed further light on the problem.

Some of the ideas used here may be helpful in defining areas to be investigated. For instance, it is reasonable to assume that the range of fields over which the distribution function $G(H_s)$ has an appreciable value, is governed by the homogeneity of the material. This is reasonable because the starting field should depend on the local properties of the material. Consequently, the uniformity of grain structure and size might well be correlated with electrical properties through this distribution function. Other types of interpretations tying electrical properties to metallurgical properties may be inferred from the model. Of course, these must be investigated in detail before they can be accepted as physical facts.

The results given in Figs. 8 and 11 show good agreement with the model. One might naturally ask what other materials can be represented by the model, and under what variety of conditions the model is valid for the materials tested here. The first question has a partial answer in some indirect evidence published about magnetic-amplifier performance over a supply frequency range.[22]

The answer to the second question is covered in some detail in a companion paper by Lovell.[24] Suffice it to say that a key assumption in the analysis was that the applied field C does not vary appreciably. This assumption was made to allow the assumption of a uniform distribution of domain walls which remains uniform throughout the flux reversal. If, for instance, the field is increased abruptly during the resetting process, new domain walls will, no doubt, be formed and their motion with their resulting $g(r)$ functions must be added, starting at a time $t>0$ in the switching process. It is doubtful that the present model would describe such a process; how well it does is discussed by Lovell.[24]

In spite of all doubts about the model, it can be useful in magnetic-amplifier analysis to improve the quality of a design procedure to one that does not require a cut-and-try process using several breadboard models to arrive at a final design.

References

1. Self-Saturation in Magnetic Amplifiers W. J. Dornhoefer. *AIEE Transactions*, vol. 68 pt. II, 1949, pp. 835–50.

2. Determination of Steady-State Performance of Self-Saturating Magnetic Amplifiers, E. J. Smith. *Ibid.*, vol. 69, pt. II, 1950, pp. 1309–17.

3. Predetermination of Control Characteristics of Half-Wave Self-Saturated Magnetic Amplifiers, H. Lehman. *Ibid.*, vol. 70, pt. II, 1951, pp. 2097–2103.

4. Magnetic Characteristics Pertinent to the Operation of Cores in Self-Saturating Magnetic Amplifiers, R. W. Roberts. *Ibid.*, pt. I (*Communication and Electronics*), vol. 73, 1954 (Jan. 1955 section), pp. 682–90.

5. Dynamic Hysteresis Loops of Several Core Materials Employed in Magnetic Amplifiers, H. W. Lord. *Ibid.*, vol. 72, Mar. 1953, pp. 85–88.

6. The Influence of Magnetic Amplifier Circuitry Upon the Operating Hysteresis Loops, H. W. Lord. *Ibid.*, 1953 (Jan. 1954 section), pp. 721–28.

7. Low-Impedance Operational Characteristics of Toroidal Cores Used in Magnetic Amplifiers, C. L. Boyajian. *Ibid.*, vol. 75, July 1956, pp. 307–13.

8. Flux Resetting Characteristics of Several Magnetic Materials, H. Huhta. *Ibid.*, vol. 73, May 1954, pp. 111–14.

9. Flux Reversal in Magnetic Amplifier Cores, F. J. Friedlander. *Ibid.*, vol. 75, July 1956, pp. 268–78.

10. Dynamic Core Behavior and Magnetic Amplifier Performance, L. A. Finzi, D. L. Critchlow. *Ibid.*, vol. 76, May 1957, pp. 229–40.

11. Modern Magnetism, L. F. Bates. Cambridge University Press, London, England, 1951, pp. 281–84.

12. Physical Theory of Ferromagnetic Domains, C. Kittel. *Reviews of Modern Physics*, New York, N. Y., vol. 21, no. 4, Oct. 1949, pp. 541–82.

13. Magnetic Domains, H. J. Williams. *Bell Laboratories Record*, New York, N. Y., vol. 30, no. 10, Oct. 1952, pp. 385–96.

14. A Nucleation Problem in Ferromagnetism, L. J. Dijkstra. *Thermodynamics in Physical Metallurgy*, Cleveland, Ohio, 1950, pp. 271–81.

15. A Theory of Domain Creation and Coercive Force in Polycrystalline Ferromagnetics, J. B. Goodenough. *Physical Review*, New York, N. Y., vol. 95, no. 4, Aug. 15, 1954, pp. 917–32.

16. On the Theory of the Dispersion of Magnetic Permeability in Ferromagnetic Bodies, L. Landau, E. Lifshitz. *Physikalische Zeitschrift der Sowjetunion*, Kharkov, U.S.S.R., vol. 8, 1935, pp. 1953–69.

17. Studies of the Propagation Velocity of a Ferromagnetic Domain Boundary, H. J. Williams, W. Shockley, C. Kittel. *Physical Review*, vol. 80, no. 6, Dec. 15, 1950, pp. 1090–94.

18. Magnetic Materials for Digital-Computer Components. I. A Theory of Flux Reversal in Polycrystalline Ferromagnetics, N. Menyuk, J. B. Goodenough. *Journal of Applied Physics*, New York, N. Y., vol. 26, no. 1, Jan. 1955, pp. 8–18.

19. Ferromagnetism at Very High Frequencies. III. Two Mechanisms of Dispersion in a Ferrite, G. F. Rado, R. W. Wright, W. H. Emerson. *Physical Review*, vol. 80, no. 2, Oct. 15, 1950, pp. 273–80.

20. Magnetic Materials for Digital-Computer Components. II. Magnetic Characteristics of Ultra-Thin Molybdenum Permalloy Cores, N. Menyuk. *Journal of Applied Physics*, vol. 26, no. 6, June 1955, pp. 692–97.

21. Magnetic Amplifier Analysis Using a Generalized Model for the Saturable Reactor Core, H. H. Woodson. *WADC Technical note 56-198*, Servomechanisms Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., May 1956.

22. Analysis of Magnetic Amplifiers Using a Resistive Reactor-Model, S. Frankenthal. *AIEE Transactions*, see pp. 849–56 of this issue.

23. A Magnetic Amplifier Analysis Using Intrinsic Core Flux Behavior, M. L. Shooman. *M.S. Thesis*, Massachusetts Institute of Technology, Aug. 1956.

24. Two-Step Switching of Saturable Reactors, B. W. Lovell. *AIEE Transactions*, see pp. 895–900 of this issue.

◆

Discussion

F. J. Friedlaender (Purdue University, Lafayette, Ind.): Dr. Woodson presents an interesting model for the reversal of flux in certain core materials. A comparison between his model and those suggested by other authors[1-3] would seem worth while. Dr. Woodson assumes that flux reversal initiates as a volume process in his cores, whereas it has been shown previously[1,2,4] that in nickel iron cores flux reversal initiates at the surface.

If flux reversal initiates at the surface, eddy currents introduce a position-dependent damping factor (eddy currents cannot be neglected in the 2-mil tape cores commonly used in magnetic amplifiers at 60 cycles per second, or in thinner tapes at higher frequencies). Dr. Woodson assumes that damping is not a function of the domain wall position and therefore does not express explicitly the resistivity of the core material. How do such factors as the resistivity, the spin relaxation damping constant, and the thickness of the cores enter into the describing functions?

Perhaps the author's model is to be considered merely as an elegant mathematical artifice for the analysis of magnetic amplifiers, in which case it is not clear how the model may be used to handle such questions as the influence of the grain size; this type of question requires a model which is more closely based on the phenomenological processes.

Incidentally, the grain size does not seem to affect the process of flux reversal significantly. This statement is based on recent experiments that we have performed on grain oriented 50% Ni-Fe tape cores with rather uniform grain sizes varying from 10^{-3} mm to larger than 1 mm.

References

1. Influence of Pulsed Magnetic Fields on the Reversal of Magnetization in Square Loop Metallic Tapes, D. S. Rodbell, C. P. Bean. *Journal of Applied Physics*, New York, N. Y., vol. 26, no. 11, Nov. 1955, pp. 1318–23.

2. See reference 9 of the paper.

3. Terminal Properties of Magnetic Cores, T. C. Chen, A. Papoulis. *Proceedings*, Institute of Radio Engineers, New York, N. Y., May 1958, p. 839.

4. Kinetics of Magnetization in Some Square Loop Magnetic Tapes, C. P. Bean, D. S. Rodbell. *Journal of Applied Physics*, vol. 26, no. 1, Jan. 1955, pp. 124–25.

R. C. Barker (Yale University, New Haven, Conn.): In the first sentence of his "Results and Conclusions" the author has made an important statement. This discussion is designed to enlarge on one or two of the points which may be responsible for the author's apprehension.

The paper is divided into two sections: the first consists in some assertions about the magnetization process in metallic tapes; the second has to do with modification and corroboration of the popular inverse-switching-time curves, as summarized in Figs. 6, 8, and 11. The question of the connection between the two may be approached by examining the use made by the author of this analytical description of the ferromagnetic process. He examines flux switching under three conditions in terms of the two functions $G(H_s)$ and $g(r)$, introduced from the first part of the paper. The first switching condition results in equation 10, which uses neither function, and is readily obtained by extending equations 1 and 2. In the second condition of switching, the choice of a power series for $G(H_s)$ results in equation 15, which is a power series in $(H-H_s)$. This amounts to the recognition that the curves of Fig. 8 rectangular paper are not linear. Standard procedure in such cases is to represent the curve by a few terms of a Taylor series expansion. Thus, a relationship similar to that of equation 15 could be obtained by again extending the meaning of equations 1 and 2. This would be entirely equivalent to the author's procedure, because he has chosen the term in the second power of $(H-H_s)$ solely on the basis of the experimental fit. These comments also extend in large part to the third condition of switching.

Thus it would seem that the second portion of the paper could stand on its own merit since it does not depend in any crucial sense upon the existence or the particular form of the functions $G(H_s)$ and $g(r)$. Certainly, to use the experimental corroboration of the equations to "prove" the validity of his assumptions and conjectures on the process of ferromagnetism would be to commit the well-known "fallacy of affirming the consequent."

With regard to the first portion of the paper, the author has been careful to warn that there are many speculative assumptions made. A debate of the validity and consistency of these assumptions would probably be beyond the author's intended scope. But assumptions 2 and 3 deserve comment. With regard to 2, the nucleation field for a domain in a pure crystal is actually larger than the field required for irreversible domain-wall motion. This fact is directly responsible for the tendency toward squareness of the hysteresis loop mentioned in assumption 3. However, as the material is brought out of saturation, the large number of independent domains and the presence of impurities does result in the fact that the field required for the nucleation and initial growth of domains is less than the "starting field" required for further irreversible growth. And the presence of metallurgical irregularities in polycrystalline materials implies further that there is some distribution of these starting fields over the wall elements about the various nucleation sites. This is true whether the domains are nucleated uniformly throughout the volume or only at the surfaces. The result is that the d-c hysteresis loop definitely has rounded coroners in polycrystalline tapes, as the measurement of a d-c hysteresis loop confirms. In this portion of the loop, the domains are changing in size and shape about the nucleating sites, and are combining to form fewer domains of smaller total wall energy. Would the author state his assumptions on the size, shape, and number of domains throughout the flux reversal process? The use of equation 2 carries with it some implications in this regard.

The author's remarks on the limited state of knowledge in this field are sincerely appreciated. Advances in this respect would undoubtedly increase our understanding of the physics of materials and would augment our creative potential as engineers. But when we consider the task of making quantitative calculations of electric circuit performance, it is interesting to contemplate the question of how we would specifically make use of a complete and detailed theory when experimental information such as that contained in Figs. 6, 8, and 11 is available to us.

H. H. Woodson: The interesting and informative discussions by two well-known workers in this field are very much appreciated. Both discussers apparently agree with the first paragraph of the "Results and Conclusions" of the paper.

I cannot fully agree or disagree with Dr. Friedlander that nucleation initiates as a surface process. I am not sure that his references prove this. For instance, his reference 1 states, "An alloy or 50% Ni, or 50% Fe can be made into a tape having a rectangular hysteresis loop by appropriate cold working and heat treating such that a strong texture is developed with the 100 plane in the plane of the tape and the 100 direction along the length of the tape. A sample of the type did not show the simple behavior previously described. Instead it appears that domains are nucleated at many places within the tape." In his reference 2, Fig. 13 is supposed to prove that nucleation is a surface process; however, the percentage spread in his points is a little too great to be completely conclusive. On the other hand, if Dr. Friedlaender has further experimental evidence of surface nucleation, I am sure it would be welcome in the literature.

I should point out that a model of the type described in this paper can equally well be adapted to the assumption of surface nucleation. The essential feature of this model is that all domains are assumed to have identical properties except for the starting field. I agree with Dr. Friedlaender that factors such as resistivity, spin relaxation damping constant, and tape thickness affect the properties of the cores; but my model does not include these effects explicitly. Further study is necessary to determine how the constants of the model are affected by material properties.

With respect to the model being an artifice for the analysis of magnetic amplifiers, Dr. Friedlaender is essentially correct. In the past, models have been so mathematically complex that they could not possibly be applied to any engineering design problem. My object in this work was to devise a model which could be used to design magnetic amplifiers and at the same time had more connection with ferromagnetic theory then previous simplified models. Along this same line, Dr. Barker's final comment indicates that perhaps a model is not really needed. Without any model, experimental information is useful only for the conditions under which it was obtained, thus a vast amount of experimental information may be necessary for a wide variety of conditions. With a reasonable model, the amount of necessary experimental information is reduced.

There were no assumptions made about the sizes, shapes, or number of domains beyond the assumption that the number is sufficiently large (and thus the sizes sufficiently small) that integration instead of summation is valid over the distribution of domains.

Theory of 3-Phase Bridge Magnetic Amplifier

I—Steady State

H. F. STORM
MEMBER AIEE

C. W. FLAIRTY
ASSOCIATE MEMBER AIEE

THE primary object of this paper is to develop a theory for the 3-phase bridge magnetic amplifier. A further object is to describe the utilization of this theory to arrive at design and performance equations which are sufficiently simple and accurate for engineering purposes. Part I deals with the steady-state operation, and Part II (which is planned for future publication) with the transient performance. Because of space limitations, the full description of the circuit components has been relegated to Part II.

With the advent of germanium and silicon rectifiers of large current capabilities, magnetic amplifiers are being used to handle power in increasing blocks. For instance, variable-speed motor drives are now available for motors of 100 horsepower and more, and power supplies for electrochemical and other applications are reaching into the 1,000-kw area.

Although the 3-phase amplifier has more parts than the single-phase version, the 3-phase connection has the following major advantages:

1. Higher utilization of magnetic and rectifier components.
2. Even distribution of load over all three phases of the supply system.
3. Smaller ripple.

For output powers of 1 kw or more, these advantages usually outweigh the drawback of requiring more parts for the 3-phase amplifier. Publications on 3-phase magnetic amplifiers are listed in references 1 through 15.

A. Circuit Connections and Supply System

The circuit under consideration is shown in Fig. 1(A); it is called a 3-phase

---

Paper 59-1060, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-IRE Nonlinear Magnetics and Magnetic Amplifiers Conference, Washington, D. C., September 23–25, 1959. Manuscript submitted May 1, 1959; made available for printing July 20, 1959.

H. F. STORM and C. W. FLAIRTY are with the General Electric Company, Schenectady, N. Y.

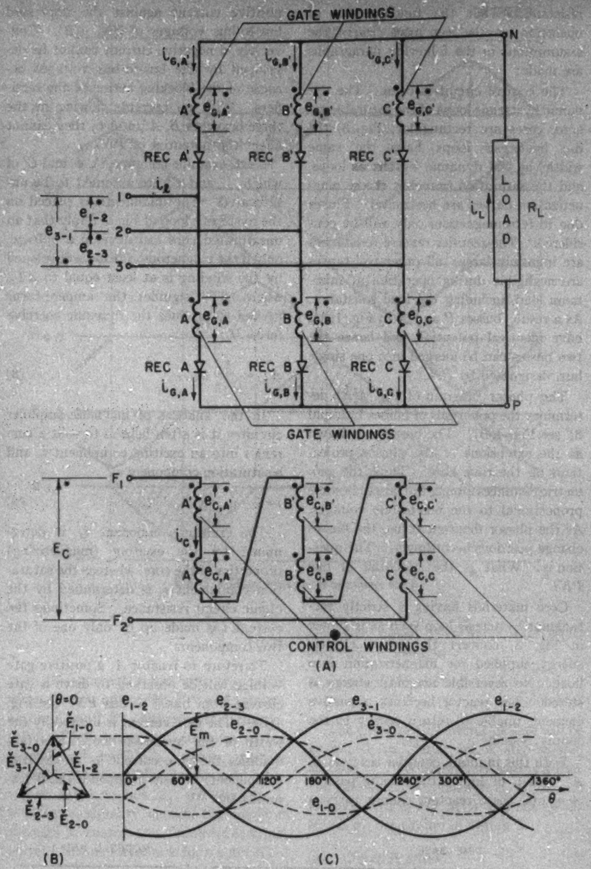
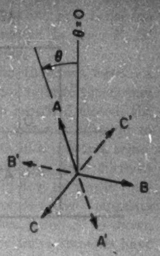

GATE WINDINGS
N
$i_{G,A'}$
$i_{G,B'}$
$i_{G,C'}$
A'
B'
C'
$e_{G,A'}$
$e_{G,B'}$
$e_{G,C'}$
REC A'
REC B'
REC C'
i_ℓ
1
2
3
e_{1-2}
e_{2-3}
e_{3-1}
i_L
LOAD
R_L
A
B
C
$e_{G,A}$
$e_{G,B}$
$e_{G,C}$
REC A
REC B
REC C
$i_{G,A}$
$i_{G,B}$
$i_{G,C}$
P
GATE WINDINGS
F_1
F_2
E_C
i_C
$e_{C,A'}$
$e_{C,B'}$
$e_{C,C'}$
$e_{C,A}$
$e_{C,B}$
$e_{C,C}$
CONTROL WINDINGS
(A)
θ=0
$\check{E}_{1-0}$
$\check{E}_{3-0}$
$\check{E}_{3-1}$
$\check{E}_{1-2}$
$\check{E}_{2-3}$
$\check{E}_{2-0}$
e_{1-2}
e_{2-3}
e_{3-1}
e_{1-2}
E_m
e_{2-0}
e_{3-0}
e_{1-0}
0°
60°
120°
180°
240°
300°
360°
θ
(B)
(C)

Fig. 1. Three-phase bridge amplifier

A—Circuit diagram
B—Phasor diagram of supply voltages. Accented E's indicate phasors
C—Sinusoidal representation of supply voltages

Fig. 2. Phase sequence diagram



bridge magnetic amplifier. The amplifier consists of six identical saturable reactors (A, C', ..., B') and six identical rectifiers (REC A, REC C', ..., REC B'). The load is a pure resistance.

Terminals 1, 2, and 3 are connected to a symmetrical 3-phase supply system, having the star voltages e_{1-0}, e_{2-0}, e_{3-0}.

$$e_{1-0}=\frac{E_m}{\sqrt{3}}\cos\theta \tag{1A}$$

$$e_{2-0}=\frac{E_m}{\sqrt{3}}\cos\left(\theta-\frac{2\pi}{3}\right) \tag{1B}$$

$$e_{3-0}=\frac{E_m}{\sqrt{3}}\cos\left(\theta-\frac{4\pi}{3}\right) \tag{1C}$$

with

$$\theta=\omega t \tag{2}$$

and

$$\omega=2\pi f \tag{3}$$

The line-to-line voltages e_{1-2}, e_{2-3}, e_{3-1} are defined by equations 4

$$e_{1-2}=e_{1-0}-e_{2-0} \tag{4A}$$

$$e_{2-3}=e_{2-0}-e_{3-0} \tag{4B}$$

$$e_{3-1}=e_{3-0}-e_{1-0} \tag{4C}$$

and, when substituted in equations 1, result in

$$e_{1-2}=E_m\sin\left(\theta+\frac{2\pi}{3}\right) \tag{5A}$$

$$e_{2-3}=E_m\sin\theta \tag{5B}$$

$$e_{3-1}=E_m\sin\left(\theta-\frac{2\pi}{3}\right) \tag{5C}$$

The supply voltages are shown in Fig. 1(B) in phasor form and in Fig. 1(C) in sinusoidal form. When using the phasor diagram the instantaneous values of the voltages are obtained by rotating the entire phasor diagram counterclockwise and by taking the projections of the phasors on the stationary base line, indicated by $\theta=0$.

The dots in Fig. 1 signify winding polarities. The voltage across a winding is defined as positive when the terminal with the dot is more positive than the other terminal. A current is called positive when it flows in the direction of the arrow. The magnetomotive force is called positive when derived from a positive current, and the core fluxes are called positive when derived from positive magnetomotive forces.

The symmetries of the circuit configuration and the supply voltages suggest that whatever occurs in one branch will be repeated sequentially in certain other branches. If a period of 360° (degrees) is assumed, and identical waveshapes in all branches, phase-displaced sequentially by 60°, the sequence can be readily visualized from the phasor diagram of Fig. 2. Suppose an occurrence (such as a gate voltage or gate current) in branch A is called $f_A(\theta)$; then corresponding occurrences in the other branches are determined by

$$f_{C'}(\theta)=f_A(\theta-60°) \tag{6A}$$

$$f_B(\theta)=f_A(\theta-120°) \tag{6B}$$

$$f_{A'}(\theta)=f_A(\theta-180°) \tag{6C}$$

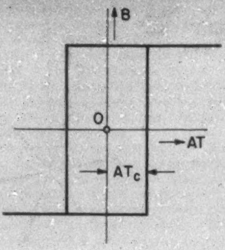


Fig. 3. Simplified B-H loops of cores. Dynamic and d-c loops requiring the same coercive ampere-turns AT_c

$$f_C(\theta)=f_A(\theta-240^\circ) \tag{6D}$$

$$f_B{}'(\theta)=f_A(\theta-300^\circ) \tag{6E}$$

Similarly, if during one and the same interval of 60° the events are known in all branches, the events in any one branch can be synthesized for the full 360° period:

$$f_C{}'(\theta)\Big|_{\theta_1}^{\theta_1+60^\circ}=f_A(\theta-60^\circ)\Big|_{\theta_1}^{\theta_1+60^\circ}=f_A(\theta)\Big|_{\theta_1-60^\circ}^{\theta_1} \tag{7A}$$

$$f_B(\theta)\Big|_{\theta_1}^{\theta_1+60^\circ}=f_A(\theta-120^\circ)\Big|_{\theta_1}^{\theta_1+60^\circ}=f_A(\theta)\Big|_{\theta_1-120^\circ}^{\theta_1-60^\circ} \tag{7B}$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$f_B{}'(\theta)\Big|_{\theta_1}^{\theta_1+60^\circ}=f_A(\theta-300^\circ)\Big|_{\theta_1}^{\theta_1+60^\circ}=f_A(\theta)\Big|_{\theta_1-300^\circ}^{\theta_1-240^\circ} \tag{7E}$$

The predominant mode of operation of a 3-phase bridge amplifier has a period of 360°, identical waveshapes in all branches, and phase displaced sequentially by 60°. Other modes are possible, e.g., in underexcitation, where the period can be 720°, and where the waveshapes in saturable reactors A, B, and C are different from the waveshapes in saturable reactors A', B', and C'; a future paper is intended to deal with these modes. In this paper the mode expressed by equations 6 and 7 will be postulated throughout.

B. Operation at Minimum Load

Minimum load (current) is the smallest value to which the current in the load resistance can be reduced by the application of control current. (The minimum load point is designated by M in Figs. 9 and 12.)

1. Control Circuit Open

At the minimum load point, the load current is limited by the unsaturated gate reactance connected in series with the load resistance. That the fundamentals of operation may be seen more clearly, the assumptions in the following paragraphs are made:

The control circuit is open. The dynamic hysteresis loops of the saturable reactor cores are rectangular, Fig. 3, the d-c hysteresis loops have the same widths as the dynamic hysteresis loops, and the saturation branches of the magnetization curves are horizontal. Fluxes due to ferromagnetism only will be considered. The rectifier reverse resistances are infinitely large; all other resistances are negligible during operation at minimum load, including the load resistance. As a result, busses P and N in Fig. 1(A), have identical potential, and hence the two busses can be merged into one single bus, designated by PN.

The phasor diagram of Fig. 4(A) determines the potentials of busses 1, 2, and 3; see Fig. 4(B). The busses are drawn as the extensions of the phasor projections on the time base. Thus, the geometric distances among the busses become proportional to the bus-to-bus voltages. As the phasor diagram turns, the busses change positions accordingly. The question is: What is the potential of bus PN?

Core materials having a strictly rectangular hysteresis loop such as is shown in Fig. 3, convert the entire electric energy supplied for magnetization into heat. No reversible magnetic energy is stored. The reactor becomes a passive element, unable to return energy to the supply system.

With this in mind, consider reactors B, A', and C of Fig. 4(B). Being passive elements, these reactors cannot carry a positive current against the impressed bus-to-bus voltages of Fig. 4(B). Conversely, a negative current cannot be developed by the bus-to-bus voltages because of the blocking action of the rectifiers. With no currents flowing in the three branches B, A', and C, they cannot affect the potential of PN.

Next, examine reactors B', A, and C' of which A, and C' are assumed to be unsaturated. The qualifications placed on the hysteresis loop of Fig. 3 imply that an unsaturated core can develop a voltage, only if the magnetomotive force developed by the winding is at least equal to AT_c, where AT_c signifies the ampere-turns needed to produce the dynamic coercive force H_c:

$$AT_c=\frac{1}{0.4\pi}H_c l_{Fe} \tag{8}$$

In the analysis of magnetic-amplifier circuitry it is often helpful to split a current i into an exciting component i_x and a saturation component i_s:

$$i=i_x+i_s \tag{9}$$

The exciting component i_x is determined by the exciting (magnetizing) properties of the core, whereas the saturation component i_s is determined by the ohmic circuit resistance. Sometimes the current i is made up by only one of the two components.

To return to reactor A, a positive gate voltage will be observed to drive a gate current from bus 1 to bus PN; see Fig. 4(B). The gate current is limited by the width of the hysteresis loop; this fact qualifies the gate current as an exciting component, whose magnitude is determined by

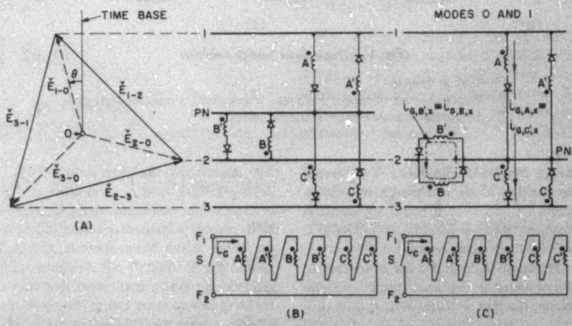


Fig. 4. Voltage and circuit configurations at minimum load

A—Phasor diagram
B—Configurations with bus PN tentative
C—Configurations with bus PN final. Accented E's indicate phasors

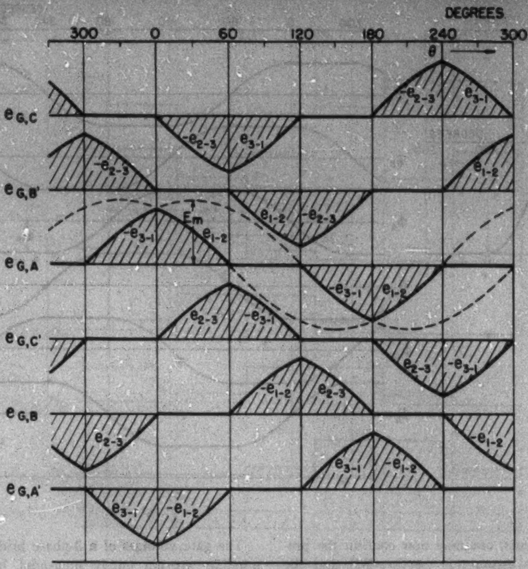


Fig. 5. Gate voltages at minimum load

$$i_{G,A,x}=\frac{AT_c}{N_G} \quad (10A)$$

If this current splits into two parts when flowing out of bus *PN* and into reactors *B'* and *C'*, neither reactor can develop AT_c ampere-turns, and hence neither reactor can develop a voltage. Consequently, the bus *PN* will drop until it becomes equal to bus 2; see Fig. 4(C). When this equality occurs, current flow through *B'* ceases, and the entire exciting current becomes available for reactor *C'*:

$$i_{G,C',x}=i_{G,A,x}=\frac{AT_c}{N_G} \quad (10B)$$

Thus, reactor *C'* can now sustain the voltage between busses *PN* and 3. By the same reasoning, one deduces that the potential of bus *PN* cannot be lower than the potential of bus 2. Hence, the answer to the question posed above is: The potentials of busses *PN* and 2 are identical for an angle *θ* as shown in Fig. 4(A). More generally, the potential of bus *PN* is identical with the smallest of the three star potentials, when measured in terms of absolute magnitudes.

2. Control Circuit Closed

The control circuit, whose resistance is assumed zero ($R_C=0$) is now being closed by means of switch *S*, Figs. 4 (B) and (C), resulting in

$$e_{C,A}+e_{C,C'}+e_{C,B}+e_{C,A'}+e_{C,C}+e_{C,B'}=0 \quad (11)$$

The effect of this circuit closure will be examined in detail by a step-by-step method. It is tentatively assumed that the gate voltages of reactors *A*, *C'*, *B*, and *B'*, as shown in Fig. 4(C) will not be affected by the closure of switch *S*.

(a) CORES A AND C'

Cores *A* and *C'* are assumed unsaturated and therefore the gate voltages will be transformed into the control circuit:

$$e_{C,A}=e_{G,A}\frac{N_C}{N_G} \quad (12A)$$

$$e_{C,C'}=e_{G,C'}\frac{N_C}{N_G} \quad (12B)$$

The voltages, $e_{C,A}$ and $e_{C,C'}$ will attempt to drive a current through the control circuit. The current flowing in the control circuit is called the "control current."

(b). CORES B, B'

Suppose these two cores are in negative saturation prior to the closure of switch *S*. The introduction of a negative control current (by closing switch *S*) cannot drive the cores more negative, and hence no voltages will be induced in cores *B*, *B'* by the control current.

Now, assume that the cores were not in negative saturation prior to the closure of switch *S*. With switch *S* closed, the negative control current may attempt to drive the core fluxes in a negative sense. If this happens, positive gate currents $i_{G,B,x}$ and $i_{G,B',x}$ will flow by transformation. These currents can circulate through the branches *B* and *B'* as shown in Fig. 4(C). Since these branches have no resistance in the forward direction (by previous assumption), the gate voltages $e_{G,B}$ and $e_{G,B'}$ remain at zero; hence the control voltages $e_{C,B}$ and $e_{C,B'}$ also remain at zero:

$$e_{G,B}=0 \quad (13A)$$

$$e_{G,B'}=0 \quad (13B)$$

$$e_{C,B}=0 \quad (13C)$$

$$e_{C,B'}=0 \quad (13D)$$

One concludes that the bus voltages shown in Fig. 4(C) remain undisturbed by the flow of control current, no matter what the state of magnetization of cores *B*, *B'*. If the control current should induce any gate currents in *B*, *B'*, these currents form theoretically a short-circuit loop, and therefore no impedance is encountered by the control current in reactors *B*, *B'*.

(c). CORES A', C

The sum of the control voltages $e_{C,A'}$ and $e_{C,C'}$ can be determined from equations 11, 13(C), and 13(D)

$$e_{C,A'}+e_{C,C}=-(e_{C,A}+e_{C,C'}) \quad (14A)$$

Transforming these voltages to the gate side yields

$$e_{G,A'}+e_{G,C}=-(e_{G,A}+e_{G,C'}) \quad (14B)$$

The forward voltages induced in gate windings *A'* and *C* cannot exceed the voltages on gates *A* and *C'* respectively:

$$-e_{G,A'}\le e_{G,A} \quad (15A)$$

$$-e_{G,C}\le e_{G,C'} \quad (15B)$$

In view of equation 14(B), the solutions are

$$e_{G,A'}=-e_{G,A} \quad (16A)$$

$$e_{G,C}=-e_{G,C'} \quad (16B)$$

and by transformation

$$e_{C,A'}=-e_{C,A} \quad (16C)$$

$$e_{C,C}=-e_{C,C'} \quad (16D)$$

In conjunction with the control current, the voltages $e_{C,A'}$, $e_{C,C}$ reduce the core fluxes in cores *A'* and *C* by the same

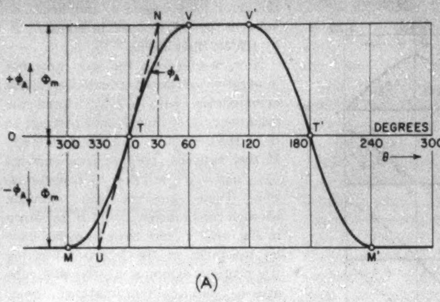
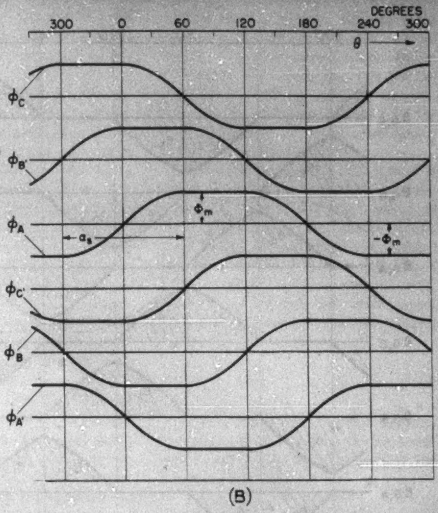


Fig. 6. Core fluxes at minimum load

A—Fluxes of core A
B—All core fluxes



amounts by which the core fluxes in cores A and C' are increased (by the supply voltages).

(d). Gate and Control Currents

As stated in section B2(b), no impedance is encountered by the control current passing through the reactors B, B'. The control current, however, will experience a limitation by the resetting cores A', C. These cores are resetting against the line-to-line voltages, shown in equations 16(A), and (B), and thereby hold their excitation $i_C N_C$ to the width of the dynamic hysteresis loop:

$$i_C = -\frac{AT_c}{N_C} \tag{17}$$

Prior to the closing of switch S, the gates A and C' were magnetized by the gate currents only, resulting in AT_c ampere-turns for each core; by closing switch S, negative ampere-turns have been added by the control ampere-turns; see equation 17. The net magnetomotive forces $\mathfrak{F}_A$, $\mathfrak{F}_C$ are

$$\mathfrak{F}_A = i_{G,A,x} N_G + i_C N_C \tag{18A}$$

$$\mathfrak{F}_{C'} = i_{G,C',x} N_G + i_C N_C \tag{18B}$$

To maintain the gate voltages $e_{G,A}$, $e_{G,C'}$, the net magnetomotive forces must produce AT_c ampere-turns:

$$\mathfrak{F}_A = \mathfrak{F}_{C'} = AT_c \tag{19}$$

Substituting equations 17 in equations 18, one obtains for the gate currents

$$i_{G,A,x} = i_{G,C',x} = \frac{2AT_c}{N_G} \tag{20}$$

One recognizes the doubling of gate currents $i_{G,A,x}$, $i_{G,C',x}$ (equations 10) as a result of closing switch S.

Having accounted for all voltages and currents, one may now confirm the previous assumption (section B2) that the closure of switch S will not disturb the gate voltages of A, C', B, B' as shown in Fig. 4(C) for open switch.

(e) Voltages

The foregoing reasoning permits the calculation of all circuit voltages. Consider again the interval $0 \leq \theta \leq 60°$. During this interval the star voltage e_{2-0}, taken absolutely, is the smallest of the three star voltages shown in Figs. 1(C) and 4(A); therefore the potential of bus PN becomes identical with the potential of bus 2; see section B1:

$$e_{G,B} = 0 \tag{13A}$$

$$e_{G,B'} = 0 \tag{13B}$$

From Fig. 4(C)

$$e_{G,A} = e_{1-2} \tag{21A}$$

$$e_{G,C'} = e_{2-3} \tag{21B}$$

and from equations 16(A) and (B), and 21(A) and (B)

$$e_{G,A'} = -e_{1-2} \tag{22A}$$

$$e_{G,C} = -e_{2-3} \tag{22B}$$

Having the voltages established in all cores for one 60° interval (equations 13, 21, and 22), the events in all cores and at all times becomes available from the periodical character of the functions. Equations 7 or the phasor diagram of Fig. 2 lead to the synthesis of all functions. The results are shown in Fig. 5.

The gate voltages of a 3-phase bridge amplifier are not purely sinusoidal, but consist of portions of two sinusoidal line-to-line voltages, each portion lasting for 60°, followed by a 60° interval during which no gate voltage exists. This pattern is repeated with opposite sign. The half-period average of the gate voltage of core A follows from

$$E_{G,A} = \frac{1}{\pi}\int_0^{\pi} e_{G,A}\,d\theta = \frac{2}{\pi}\int_0^{\pi/3} e_{1-2}\,d\theta \tag{23}$$

Substituting for e_{1-2} from equations 5 yields

$$E_{G,A} = \frac{1}{\pi} E_m \tag{24}$$

It will be observed that the half-period average gate voltage, given above, of a 3-phase bridge connected amplifier is precisely one half of the half-period average of a purely sinusoidal waveshape having an amplitude of E_m.

(f). Core Fluxes

The change in core flux $\Delta\phi$ follows from Faraday's law in integral form

$$\Delta\phi = \frac{10^8}{N}\int_{t_1}^{t_2} e\,dt \tag{25}$$

The maximum flux swing $2\Phi_m$ occurs after application of gate voltage for 120° continuously. For instance, for gate A

$$2\Phi_m = \frac{10^8}{\omega N_G}\int_{\omega t=-\pi/3}^{\omega t=\pi/3} e_{G,A}\,d\theta \tag{26}$$

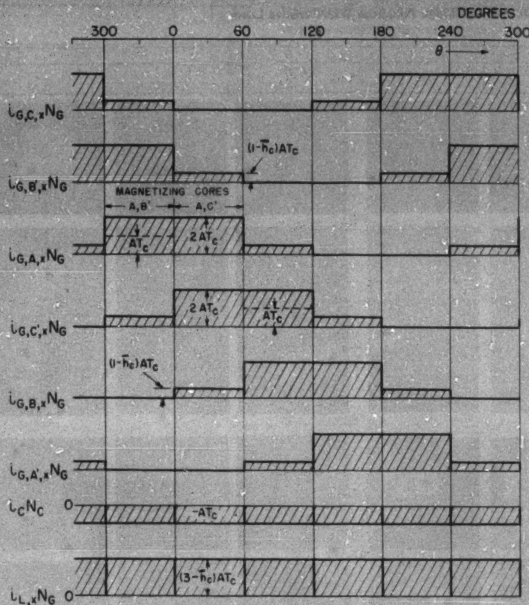


Fig. 7. Gate currents at minimum load

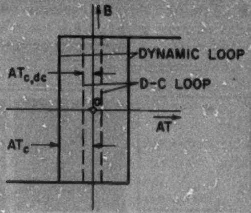


Fig. 8. Dynamic and d-c B-H loops

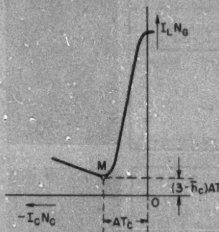


Fig. 9. Position of minimum load point M on control characteristic

From Fig. 5

$$\Phi_m=\frac{10^8}{\omega N_G}\int_{\omega t=-\pi/3}^{\omega t=0}-e_{3-1}d\theta \qquad (27)$$

Substituting for e_{3-1} from equation 5(C) and integrating, yields

$$\Phi_m=\frac{1}{2}\,\frac{10^8}{\omega N_G}\,E_m \qquad (28)$$

Assuming that the flux swings equal amounts in the positive and negative directions, the gate flux of A will be $\phi_A=+\Phi_m$ at $\theta=60°$, and $\phi_A=-\Phi_m$ at $\theta=240°$. Since no gate voltages are applied for $240°\leq\theta\leq300°$ and $60°\leq\theta\leq120°$, the core flux will remain constant:

$$\phi_A=-\Phi_m \qquad 240°\leq\theta\leq300° \qquad (29A)$$

$$\phi_A=+\Phi_m \qquad 60°\leq\theta\leq120° \qquad (29B)$$

In the remaining intervals, gate voltages are present, and the fluxes are computed from equation 25 and recast into the following form:

$$\phi_A=\frac{10^8}{N_G}\int e_{G,A}dt+C \qquad (30)$$

For the interval $300°\leq\theta\leq360°$, $e_{G,A}=-e_{3-1}$. Substituting for e_{3-1}, from equation 5(C), and integrating equation 30, yields, with the aid of equation 28,

$$\phi_A=2\Phi_m\sin(\theta-30°)+C \qquad 300°\leq\theta\leq360° \qquad (31)$$

for $\phi_A=-\Phi_m$ at $\theta=300°$, equation 29(A), yielding the integration constant C:

$$C=\Phi_m \qquad (32)$$

Substituting equation 32 into equation 31, yields for core flux ϕ_A

$$\phi_A=\Phi_m[2\sin(\theta-30°)+1] \qquad 300°\leq\theta\leq360° \qquad (33A)$$

This flux is shown in Fig. 6(A) by MTN; the part TN is merely shown to provide a complete picture of one quarter of a sine function.

For the next 60° interval the core flux ϕ_A is derived in similarity to equation 33(A):

$$\phi_A=\Phi_m[2\sin(\theta+30°)-1] \qquad 0°\leq\theta\leq60° \qquad (33B)$$

and is shown in Fig. 6(A) by UTV. TV relates to the actual core flux and UT merely completes the first portion of the sine function. The two curves MT and TV have a common tangent at point T.

Similarly, one obtains

$$\phi_A=\Phi_m[2\sin(\theta-30°)-1] \qquad 120°\leq\theta\leq180° \qquad (33C)$$

$$\phi_A=\Phi_m[2\sin(\theta+30°)+1] \qquad 180°\leq\theta\leq240° \qquad (33D)$$

as shown by curves $V'T'$ and $T'M'$ in Fig. 6(A).

Having the flux established in one core for the entire period of 360°, the core fluxes in the other cores follow in cyclic order (equations 6) and are shown in Fig. 6(B).

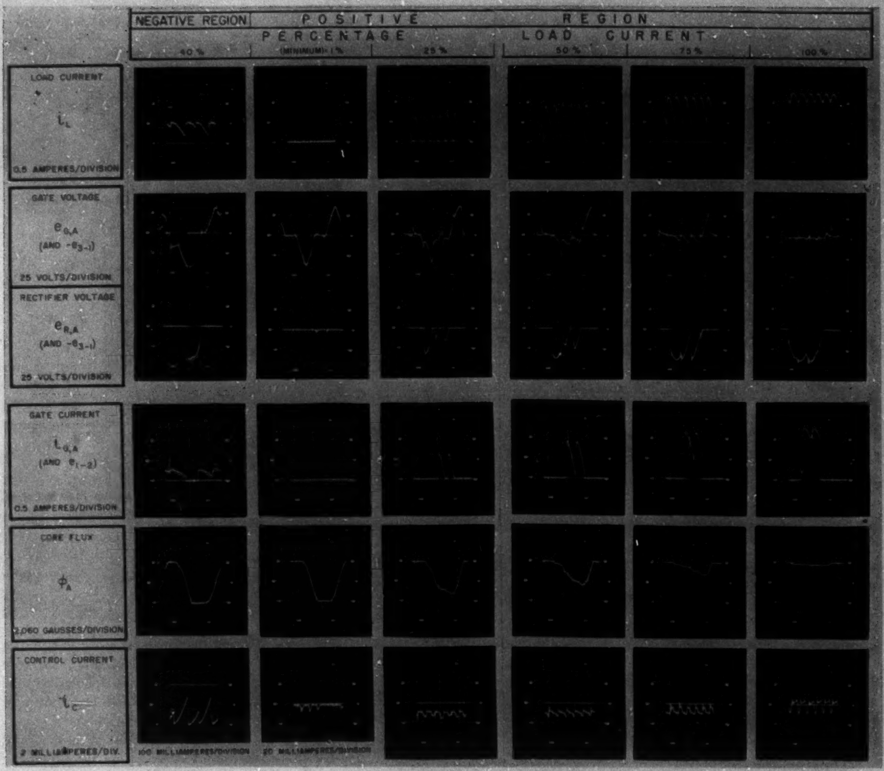
The core behavior may be summarized as follows: Two cores are increasing the flux in the positive direction (called setting), two are increasing the flux in the negative direction (called resetting), and two maintain constant flux level. Cores which are out of phase with each other by 180°, such as A A', B B', and C C', will set and reset at equal rates.

(g). DIFFERENT WIDTHS FOR D-C AND DYNAMIC HYSTERESIS LOOPS

The gate currents $i_{G,A,x}$ and $i_{G,C',x}$ were determined in section B2(d) by equation 20, for $0°\leq\theta\leq60°$, and were plotted for the above interval in the form of ampere-turns in Fig. 7.

The ampere-turns, Fig. 7, were obtained under the assumption of equal widths for the d-c and the dynamic hysteresis loops (section B1), such as in ferrites. If, however, metallic cores are used, e.g., a 50-50 Ni-Fe alloy, the d-c hysteresis loop is narrower than the dynamic hysteresis loop. This is shown

Table I. Three-Phase Bridge Amplistat With Resistive Load



in Fig. 8 where the d-c hysteresis loop is indicated by dashed lines. Of the conclusions reached in section B2, only the subsection B2(b) requires further attention because of the introduction of different widths for dynamic and d-c loops.

As in any transformer, the sum of primary ampere-turns $i_C N_C$ and secondary ampere-turns $i_{G,B',x} N_G$ equal the exciting ampere-turns $(iN)_x$:

$$i_C N_C + i_{G,B',x} N_G = (iN)_x \qquad (34)$$

and

$$i_{G,B',x} N_G = -i_C N_C + (iN)_x \qquad (35)$$

With the core fluxes in B, B' being theoretically constant, $0° \leq \theta \leq 60°$; see Fig. 6(B), the hysteresis loop, representative for zero rate of flux change, is the so-called d-c hysteresis loop; see Fig. 8. It is this loop which becomes pertinent for the transformation described by equation 35:

$$(iN)_x = -AT_{c,dc} \qquad (36)$$

Actually the resistances through which the circulating gate currents are flowing are not zero, and therefore voltage drops develop. However, these voltage drops amount in practical cases to only a small fraction of 1% based on the supply voltage E_m. The resultant flux rates are so low that the hysteresis loops on which the cores are operating are practically identical with the d-c loops.

The control ampere-turns are

$$i_C N_C = -AT_c \qquad (17)$$

By combining equations 17, 35, and 36

$$i_{G,B',x} N_G = AT_c - AT_{c,dc} \qquad (37)$$

The per-unit d-c coercive force $\bar{h}_c$ may be defined as the ratio of the d-c coercive force $H_{c,dc}$ to the dynamic coercive force H_c; then one may express $AT_{c,dc}$ in terms of $\bar{h}_c$, and the coercive ampere-turns AT_c:

$$\bar{h}_c = \frac{H_{c,dc}}{H_c} = \frac{AT_{c,dc}}{AT_c} \qquad (38)$$

One obtains from equations 37 and 38 for the gate exciting ampere-turns in B or B' $(0° \leq \theta \leq 60°)$

$$i_{G,B',x} N_G = i_{G,B,x} N_G = AT_c(1 - \bar{h}_c) \qquad (39)$$

also shown in Figs. 4(C) and 7. In the case of reactors having Deltamax cores of 4-mil thickness, and at 60 cps, $\bar{h}_c = 0.415$.

Having established all gate currents in one 60° interval, the gate currents in all cores and at all times becomes avail-

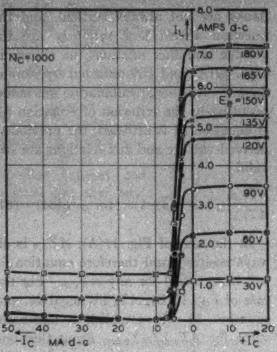


Fig. 10. Control characteristics of amplifier for various supply voltages; E_s=150 volts corresponds to normal excitation

able from the periodic character of the functions. Equations 7 or the phasor diagram of Fig 2 lead to the synthesis of all functions. The results are shown in Fig. 7.

One is now in a position to determine the load current i_L. It will be seen from Fig. 1 that gate curents $i_{G,A,x}$, $i_{G,B,x}$ flow through the load, and return to the supply as gate currents $i_{G,C',x}$, and $i_{G,B',x}$:

$$i_{L,x}=i_{G,A,x}+i_{G,B,x}\equiv i_{G,C',x}+i_{G,B',x} \quad (40)$$

From equations 20, 39, and 40 one obtains the exciting components $i_{L,x}$ of the load ampere-turns. If this analysis is restricted to normal and underexcitation, section B2(i), there is no saturation component $i_{L,s}$ of the load current at minimum load, and therefore the load current i_L consists of the exciting component $i_{L,x}$ only:

$$i_{L,x}N_G=i_LN_G=AT_c(3-\bar{h}_c) \quad (41)$$

Thus the minimum load point M (Fig. 9) which is a cardinal point of the control characteristic, has been determined in terms of the physical properties of the core; see equations 8 and 38.

$$AT_c=\frac{1}{0.4\pi}H_cl_{Fe} \quad (8)$$

As seen from Fig. 7 a direct current flows in the control winding. In the past, all forward circuit resistances were assumed zero and, therefore, once a current has been established in the control winding, no external d-c source is needed to maintain it. In a real magnetic amplifier however, the control circuit has finite resistance, and therefore an external direct voltage source E_C must be connected to terminals F_1, F_2 (Fig. 1) in order to maintain the control current (equation 17):

$$E_C=-\frac{AT_c}{N_C}R_C \quad (42)$$

(h). Reverse Rectifier Voltages

During the interval 0°≤θ≤60°, currents are flowing in rectifiers B', A, C', B (Fig. 7), and therefore no reverse voltages exist on these rectifiers.

$$e_{R,B'}=e_{R,A}=e_{R,C'}=e_{R,B}=0 \quad (43)$$

The voltage $e_{R,C}$ on rectifier C follows from the loop equation Fig. 1(A)

$$e_{R,C}+e_{G,C}+e_{R,C'}+e_{G,C'}+i_LR_L=0 \quad (44)$$

Assuming all circuit resistances as zero, and substituting from equations 16(B) and 43, one obtains

$$e_{R,C}=0 \quad (45A)$$

and, similarly

$$e_{R,A'}=0 \quad (45B)$$

The conclusion is that, with zero resistances in the circuit (except for rectifier reverse resistances), the rectifier reverse voltages are zero at the minimum load point M. This is borne out essentially in the oscillogram in Table I. The residual voltages are due to core asymmetries, also visible in the trace for the control current which should be entirely smooth. Table I also contains oscillograms for load current, gate voltage, gate current, and core flux.

(i). Various States of Excitation

With a given set of reactors, one wishes normally to obtain the smallest possible load current at minimum load, consistent with full utilization of the reactors. This condition is realized by applying a supply voltage of such a magnitude E_m, as to make the flux excursion Φ_m equal to the saturation flux Φ_s. This condition is called normal excitation.

$$\Phi_m=\Phi_s \quad (46)$$

By reducing the supply voltage, E_m, Φ_m is reduced and $\Phi_m<\Phi_s$, a condition called underexcitation. All equations derived thus far are equally applicable to normal or underexcitation. While the minimum load current i_L (equation 41) remains the same theoretically, the output voltage of the amplifier declines proportionally to the supply voltage, and hence the utilization of the reactors declines.

By increasing the supply voltage E_m, one may obtain $\Phi_m>\Phi_s$, a condition called overexcitation. In this mode of operation, the load current is no longer limited by an unsaturated gate reactance during the entire duration of each 60° interval, and therefore the load current rises at minimum load. This is shown in section E, and in Fig. 10, where normal excitation occurs for $E_s=150$ volts.

To return to normal excitation and underexcitation, the pattern of core behavior has been summarized at the end of section B2(f). In anticipation of amplifier analysis between minimum and full load, a distinction will be made for the two cores with constant flux (cores B, B', 0≤θ≤60°). When the core flux is in positive saturation, the core will be called saturated; if not in positive saturation, the core is called idle.

With these definitions in mind, reference is made to Table II. For minimum load and underexcitation, the asterisk indicates mode 0 operation, consisting of no core saturated, and of two cores setting, two resetting, and two idle, at any time. For minimum load and normal excitation, the asterisk indicates mode 1 operation: one saturated core, two setting cores, two resetting, and one idle core at any time. It will be noted that the designation number of the mode signifies the number of saturated cores.

(j). Saturation Angle α_s

As seen from Fig. 6(B), core A saturates at θ=60° when normally excited. Since

Table II. Functional Correlations

DESIGNATION OF AMPLIFIER OPERATION		SATURATION ANGLE	DESIGNATION OF MODE					
		α_s	0	1	2	3		
MINIMUM LOAD	†	NONEXISTENT	*					
	‡	120°		*				
MINIMUM TO ½ LOAD	‡	120°–60°		*	*			
½ TO FULL LOAD	‡	60°–0°			*	*		
			0	1	2	3	SATURATED	NUMBER (QUANTITY) OF CORES
			2	2	1	0	SETTING	
			2	2	3	0	RESETTING	
			2	1	0	3	IDLE	

FOOT NOTES:
- * MODE(S) IN USE
- † UNDER-EXCITATION
- ‡ NORMAL EXCITATION

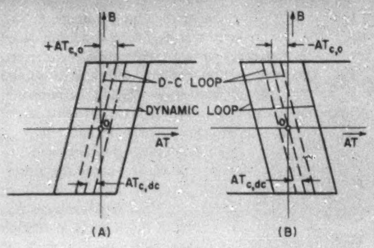


Fig. 11. B-H loops with tilted flanks

there are six cores, but only one time scale, there will be six different values of θ for the six cores. To unify the expressions, a new variable α is introduced. This variable α is counted for each core from the instant at which each core flux starts on its upswing. In this notation, the saturation angle $\alpha_s = 120°$ for the flux configuration of Fig. 6(B), and normal excitation.

(k). NORMAL EXCITATION

In view of its pre-eminence in applications, only normal excitation (equation 46) will subsequently be treated. Substituting equations 28 and 46 yields

$$\frac{1}{2}E_m = \omega N_G \Phi_s 10^{-8} \qquad (47)$$

Designating the effective rms value of the line-to-line voltage by E_s yields

$$E_s = \frac{E_m}{\sqrt{2}} \qquad (48)$$

and one obtains from equations 3, 47, and 48

$$\frac{1}{2}E_s = \pi\sqrt{2} f N_G \Phi_s 10^{-8} \qquad (49)$$

The saturation flux Φ_s is related to the effective area of core iron A_{Fe} and the saturation flux density B_s by

$$\Phi_s = A_{Fe} B_s \qquad (50)$$

Equation 49 is a fundamental design equation relating the line-to-line voltage to the physical properties of the reactor. The right-hand term of equation 49 expresses the effective voltage which can be applied to an ordinary a-c reactor, when energized from a purely sinusoidal voltage source. For the bridge-connected amplifier, however, the left-hand term of equation 49 is only $E_s/2$, and hence the line-to-line voltage of the 3-phase bridge amplifier, can be double the voltage for ordinary, sinusoidal, single-phase a-c reactor operation. For instance, if the single-phase rating of one reactor is 115 volts rms, then the same reactor can be used in a 3-phase bridge amplifier, having a line-to-line voltage of 230 volts rms, assuming the same frequency and normal excitation in both cases.

Up to this point the derivations of voltages and fluxes have been based on a rectangular hysteresis loop as shown in Fig. 3. In actual hysteresis loops, the flanks are not always vertical, and the question arises: how seriously are the voltage and flux relations, as shown in Figs. 11(A) and (B), affected by sloping flanks? A tedious examination was made, resulting in the criterion of equation 51. If equation 51 is satisfied, the previously derived voltage and flux relations are still valid

$$\frac{AT_{c,dc}}{AT_{c,0}} \leq 1-(\sqrt{3}-1)2 \quad (\text{or } -0.464) \qquad (51)$$

In the case of Fig. 11(A), $AT_{c,0}$ is always positive, and therefore equation 51 will be satisfied for any $AT_{c,0}$. For the case of Fig. 11(B), $AT_{c,0}$ is negative, and equation 51 will establish a minimum value for $AT_{c,dc}/AT_{c,0}$. From this, one concludes that the voltage and flux relations are immune to changes of the flank angles (within the limits of equation 51). It will also be observed that the waveshapes of the exciting currents in gate and control windings are sensitive to changes in the flank angles.

C. Minimum to Half-Load, $60° \leq \alpha_s \leq 120°$, Normal Excitation

"Minimum to half-load" refers to the magnitude of the load current, which will be varied in this section from its minimum value (equation 41 and M in Fig. 9) to one half of the maximum load current $I_{L,m}$ for $\alpha_s = 60°$, N in Fig. 12(B); $I_{L,m}$ follows from equation 73. This variation of load current is achieved by varying the control current, thus shifting the saturation angle α_s from 120° to 60°; the load resistance is kept constant.

As in the preceding, the analysis will be carried out within one 60° interval. It will be readily seen in the following that

Vol
Pt.
195

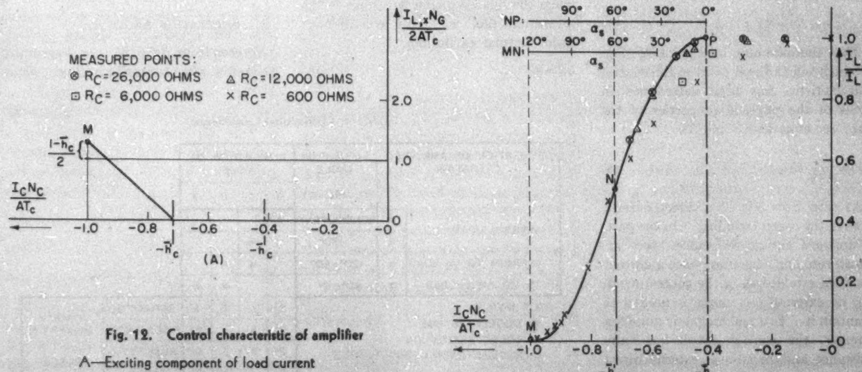


Fig. 12. Control characteristic of amplifier

A—Exciting component of load current
B—Saturation component of load current (in full line)

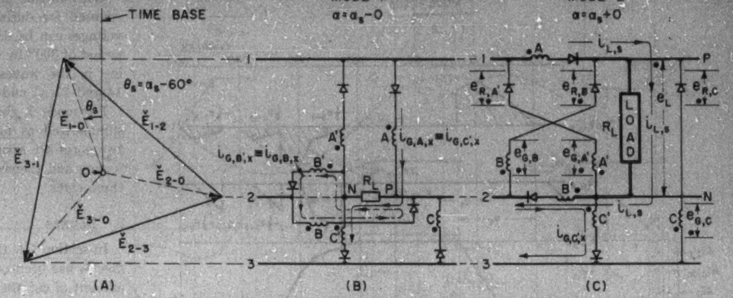


Fig. 13. Bus voltages at 25% load, $\alpha_s = 78.5°$, $\theta_s = \alpha_s - 60°$

A—Phasor diagram. Accented E's indicate phasors

B—Bus voltages and current just before saturation

C—Bus voltages and current just after saturation

two modes of operation occur in one 60° interval. For this reason the 60° interval is subdivided into two subintervals, accommodating one mode of operation for each subinterval.

1. Reactor and Rectifier Voltages

Mode 1 ($60° \leq \alpha \leq \alpha_s$)

In section B2(f), the core fluxes were $-\Phi_m$ at the start of their upswings. One may call the flux at the instant where it starts on its upswing the starting flux point Φ_1; then $\Phi_1 = -\Phi_s$ at the minimum load point.

Now, let all starting fluxes be raised by a value Φ_0, keeping Φ_0 within $-\Phi_s$ and 0. In this subinterval, the mode of excitation is precisely the same as for minimum load, normal excitation (sections B2(f) and (i), and Table II). All gate voltages are given by equations 13, 21, and 22, and are shown in Fig. 4(C). The latter figure has been redrawn in Fig. 13(B) to include the load resistance R_L, which soon will become important.

Mode 2 ($\alpha_s \leq \alpha \leq 120°$)

A new mode will be encountered in the subinterval reaching from α_s to 120°. Of the two setting (climbing) core fluxes ϕ_A, ϕ'_C, flux ϕ_A will reach saturation first. When core A saturates at $\alpha = \alpha_s$, the gate voltage $e_{G,A}$ collapses, as shown in Fig. 14. The bus P is hereby pulled up to the level of bus 1, as illustrated in Fig. 13(C). A heavy load current begins to flow, limited only by the supply voltage and the circuit resistance. The load current permits the much smaller gate exciting current $i_{G,C',x}$ to flow in the reverse direction of rectifier B'. Core C' continues to set as before equation 21(B), and induces a voltage into the control circuit equation 12(B). Since cores A and B' are saturated, $e_{C,A} = e_{C,B'} = 0$, and one obtains from equation 11

$$e_{C,B} + e_{C,A'} + e_{C,C} = -e_{C,C'} \tag{52}$$

In this mode, C' is the only core setting, and this one core is resetting the three cores B, A', and C. For equal reactors, and no currents flowing in the rectifiers B, A', and C ($\theta_s < \theta < \theta_u$ from Fig. 14):

$$e_{C,B} = e_{C,A'} = e_{C,C} = -\frac{1}{3} e_{C,C'} \quad \theta_s \leq \theta \leq \theta_u \tag{53A}$$

and the corresponding gate voltages are

$$e_{G,B} = e_{G,A'} = e_{G,C} = -\frac{1}{3} e_{G,C'} \quad \theta_s \leq \theta \leq \theta_u \tag{53B}$$

with $e_{G,C'}$ known, from equation 21(B). The rectifier voltages $e_{R,B}$, $e_{R,A'}$, and $e_{R,C}$ follow from an inspection of Fig. 13(C):

$$e_{R,B} = -e_{1-2} - e_{G,B} \tag{54A}$$

$$e_{R,A'} = -e_{1-2} - e_{G,A'} \tag{54B}$$

$$e_{R,C} = e_{3-1} - e_{G,C} \tag{54C}$$

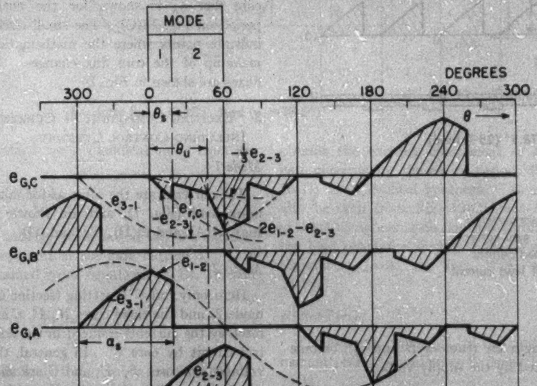


Fig. 14. Gate voltages at 25% load, $\alpha_s = 78.5°$

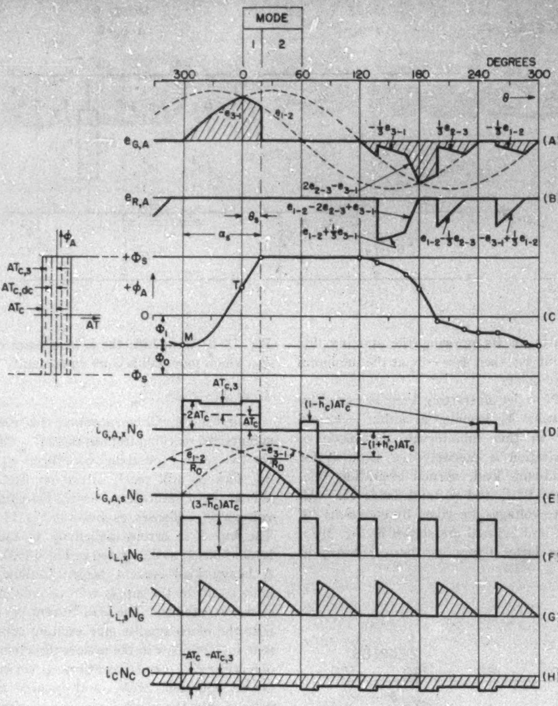


Fig. 15. Saturation angle $\alpha_s=78.5°$ (25% load)

A—Gate voltage
B—Rectifier voltage
C—Core flux
D—Exciting component of gate current
E—Saturation component of gate current
F—Exciting component of load current
G—Saturation component of load current
H—Control current

During the interval $(\theta_s \leq \theta \leq \theta_u)$ the gate voltages are given by equation 53(A). Substituting equation 53(A) into equations 54 results in the negative rectifier voltages. These voltages are called reverse voltages $e_{r,B}$, $e_{r,A'}$, $e_{r,C}$, and are shown in Fig. 14. It may be seen that reverse voltages are present for $\theta \leq \theta_u$, and that therefore no gate currents can flow, as assumed above. In fact, the angle θ_u has been defined as occurring when the rectifier reverse voltages $e_{r,B}$, $e_{r,A'}$ vanish, namely, for $e_{1-2}=1/3e_{2-3}$. This angle $\theta_u=46.1°$ when it is computed from equations 5(A) and (B).

For $\theta_u \leq \theta \leq 60°$ the reverse voltages on rectifiers B and A' vanish, and the voltages on reactors B and A' become limited by the supply voltage

$$e_{G,B}=e_{G,A'}=-e_{1-2} \quad \theta_u \leq \theta \leq 60° \tag{55A}$$

$$e_{C,B}=e_{C,A'}=-e_{1-2}\frac{N_C}{N_G} \quad \theta_u \leq \theta \leq 60° \tag{55B}$$

From equation 52

$$e_{C,C}=-e_{C,C'}-(e_{C,B}+e_{C,A'}) \tag{56}$$

From equation 21(B), 55, and 56 one obtains gate voltage $e_{G,C}$:

$$e_{G,C}=2e_{1-2}-e_{2-3} \quad \theta_u \leq \theta \leq 60° \tag{57}$$

The rectifier voltage $e_{R,C}$ is calculated from equations 54(C) and 57, and shown in Fig. 14.

With all gate and rectifier voltages accounted for during a 60° period, these voltages can be determined for an entire period of 360° by cyclic synthesis, resulting in the waveshapes shown in Figs. 14 and 15(A) and (B).

For mode 2, the core behavior can be summarized as follows, from Table II: two cores are saturated, and one core is setting and thereby resets the remaining three cores.

2. Fluxes

In section C1, the level of the starting flux Φ_1 has been raised from $-\Phi_s$ by the amount of ϕ_0; therefore core flux ϕ_A can be obtained from equation 33(A) by the addition of Φ_0:

$$\phi_A=\Phi_s[2 \sin(\theta-30°)+1]+\Phi_0 \qquad 300° \leq \theta \leq 360° \tag{58A}$$

and, from equation 33(B),

$$\phi_A=\Phi_s[2 \sin(\theta+30°)-1]+\Phi_0 \qquad 0° \leq \theta \leq 300°+\alpha_s \quad 60° \leq \alpha_s \leq 120° \tag{58B}$$

For the remaining part of the period, the flux is obtained by a step-by-step integration of equation 30, with the gate voltage $e_{G,A}$ taken from Fig. 15(A). The core flux ϕ_A is shown for the entire period on Fig. 15(C). The small circles indicate points where the mathematical make-up of the core flux changes. All fluxes are shown in Fig. 16.

3. Exciting Component of Currents Including Control Current

Mode 1

The currents are the same as for minimum load (Fig. 7), and are shown in Figs. 17(A), and 15(D), (F), and (H).

Mode 2

Here, only core C' is setting (section C1 mode 2) and the three cores B, A', C are reset by the currents induced in the control circuit by core C'. In general, the voltages on cores B, A', and C are substantially smaller than the line-to-line voltages (Fig. 14, $\theta_s \leq \theta \leq 60°$) and hence the width of the dynamic hysteresis loops used for resetting these cores will be smaller than AT_c. To appraise the effective widths of the hysteresis loops, the operation of the three resetting cores has been simulated by connecting the windings of three reactors, in series, and applying rated alternating voltage. The width of the hysteresis loop obtained in this manner is called $AT_{c,3}$. In the case of reactors having Deltamax cores of 4-mil thickness, and at 60 cps, $AT_{c,3}$ is about 0.71 AT_c. The control current is now

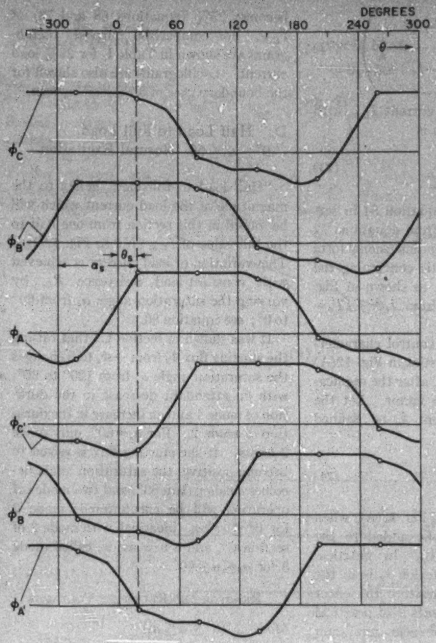


Fig. 16. Core fluxes at 25% load ($\alpha_s = 78.5°$)

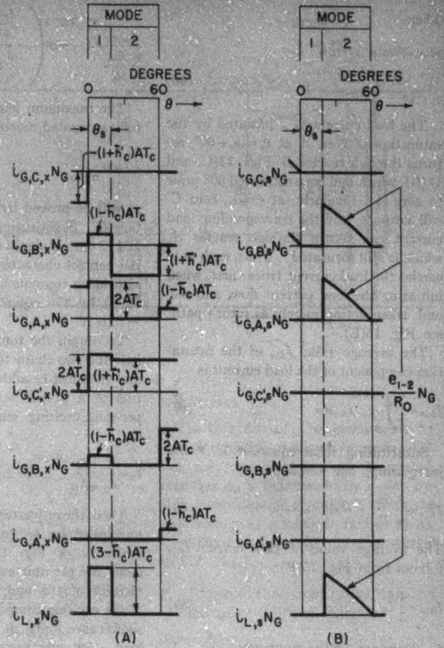

MODE
1
2
DEGREES
0
60
θ→
θ_s
$-(1+\bar{h}'_c)AT_c$
$i_{G,C,x}N_G$
$(1-\bar{h}_c)AT_c$
$i_{G,B',x}N_G$
$-(1+\bar{h}'_c)AT_c$
$2AT_c$
$(1-\bar{h}_c)AT_c$
$i_{G,A,x}N_G$
$2AT_c$
$(1+\bar{h}'_c)AT_c$
$i_{G,C',x}N_G$
$(1-\bar{h}_c)AT_c$
$2AT_c$
$i_{G,B,x}N_G$
$(1-\bar{h}_c)AT_c$
$i_{G,A',x}N_G$
$(3-\bar{h}_c)AT_c$
$i_{L,x}N_G$
(A)
MODE
1
2
DEGREES
0
60
θ→
θ_s
$i_{G,C,s}N_G$
$i_{G,B',s}N_G$
$i_{G,A,s}N_G$
$i_{G,C',s}N_G$
$\frac{e_{1-2}}{R_O}N_G$
$i_{G,B,s}N_G$
$i_{G,A',s}N_G$
$i_{L,s}N_G$
(B)

Fig. 17. Exciting and saturating components of gate currents and load currents at 25% load ($\alpha_s = 78.5°$)

obtained by a modification of equation 17:

$$i_C = -\frac{AT_{c,s}}{N_C} \qquad (59)$$

In analogy to equation 20, the exciting component of the gate current $i_{G,C',x}$ is obtained from equations 18(B), 19, and 59;

$$i_{G,C',x} = \frac{(AT_c + AT_{c,s})}{N_G} \qquad (60A)$$

Similarly to $\bar{h}_c$ in equation 38, $\bar{h}_c'$ is defined by

$$\bar{h}_c' = \frac{H_{c,s}}{H_c} = \frac{AT_{c,s}}{AT_c} \qquad (61)$$

One obtains the exciting component of the gate current C' from equations 60(A) and 61

$$i_{G,C'x}N_G = AT_c(1+\bar{h}_c') \qquad (60B)$$

The current $i_{G,C',x}$ flows through the gate B', Fig. 13(C), and therefore

$$i_{G,B',x} = -i_{G,C'x} \qquad (60C)$$

The exciting currents are plotted in Fig. 17(A). From this illustration one obtains the exciting component of the gate currents for the entire period, Fig. 15(D) by a cyclical synthesis. It will also be seen from Fig. 13(C) that the exciting components do not flow through the load resistance during mode 2, and therefore $i_{L,x}=0$; see Figs. 17(A) and 15(F).

Entire Period

The average value $I_C N_C$ of the control current can be calculated from Fig. 15(H):

$$I_C N_C = -\frac{3}{\pi} AT_c \left(\int_0^{\alpha_s - 60°} d\theta + \bar{h}_c' \int_{\alpha_s - 60°}^{\pi/3} d\theta \right) \qquad (62)$$

$$I_C N_C = -AT_c \left[\bar{h}_c' + \frac{3}{\pi}(1-\bar{h}_c')\left(\alpha_s - \frac{\pi}{3}\right) \right] \qquad (63)$$

Similarly, exciting component $I_{L,x}N_G$ of the load ampere-turns follows from Fig. 15(F):

$$I_{L,x}N_G = AT_c \times \frac{3}{\pi}\,(3-\bar{h}_c)\left(\alpha_s - \frac{\pi}{3}\right) \qquad (64)$$

4. Saturation Component of Currents

Mode 1

None.

Mode 2

The current components $i_{G,A,s}$, $i_{G,B',s}$, $i_{L,s}$ are usually several decades larger than the exciting components $i_{G,A,x}$, $i_{G,B',x}$, $i_{L,x}$. For clearer presentation the saturation and exciting components are shown on different scales in Figs. 17 and 15(D) through (G). According to equation 9, the total currents are obtained by adding the saturation and exciting components after the proper scale factors have been applied.

As seen from Fig. 13(C), the saturation component $i_{L,s}$ of the load current is limited only by the output resistance R_0. The latter consists of load resistances R_L, gate resistances $R_{G,A}$, $R_{,GB'}$ and the rectifier forward resistances $R_{f,A}$, $R_{f,B'}$. Assume that all gate resistances R_G and all rectifier forward resistances R_f are equal:

$$R_0 = R_L + 2R_G + 2R_f \qquad (65)$$

Then

$$i_{G,A,s} \equiv i_{G,B',s} \equiv i_{L,s}$$
$$= \frac{e_{1-2}}{R_0} \qquad \alpha_s - 60^\circ \leq \theta \leq 60^\circ \quad (66)$$

The load current $i_{L,s}$, initiated by the saturation of core A at $\theta_s = \alpha_s - 60^\circ$ returns through reactor B' Figs. 13(C) and 17(B), which had been saturated 60° prior to core A. Similarly, at $\theta = \alpha_s$, core C' will saturate, and the corresponding load current can return through reactor A, which is still saturated. Thus, each gate carries the load current twice: first, when initiating the load current flow, and second, when acting merely as return path; see Fig. 15(E).

The average value $I_{L,s}$ of the saturation component of the load current is

$$I_{L,s} = \frac{3}{\pi} \int_{\alpha_s}^{2\pi/3} i_{L,s} d\alpha \quad (67)$$

Substituting from equation 5(A) and integrating, one obtains

$$I_{L,s} = \frac{3}{\pi} \frac{E_m}{R_0} [1 - \sin(\alpha_s - 30^\circ)] \quad (68)$$

The average value of gate current $I_{G,s}$ follows from Fig. 15(E):

$$I_{G,s} = \frac{1}{3} I_{L,s} \quad (69)$$

5. Control Characteristics

The control characteristic of a magnetic amplifier relates the load current to the control current. By relating the exciting and saturation components of the load current to the control current, and by adding the two components, the total control characteristic is obtained.

First, the saturation angle α_s is determined from equation 63:

$$\alpha_s = -\frac{\dfrac{I_c N_c}{AT_c} + \bar{h}_c'}{\dfrac{3}{\pi}(1 - \bar{h}_c')} + \frac{\pi}{3} \qquad 60^\circ \leq \alpha_s \leq 120^\circ \quad (70)$$

The exciting component of the load current can now be obtained from equations 61 and 64. By relating the exciting load ampere-turns to the exciting gate ampere-turns $2AT_c$ for minimum load, one obtains in nondimensional form

$$\frac{I_{L,x} N_g}{2AT_c} = \left(1 + \frac{1 - \bar{h}_c}{2}\right) \frac{\left(-\dfrac{I_C N_C}{AT_c} - \bar{h}_c'\right)}{1 - \bar{h}_c'} \quad (71)$$

Equation 71 shows a linear relationship between the exciting component of the load current and the control current; see Fig. 12(A).

By substituting for α_s from equation 70 into equation 68, one obtains

$$\frac{I_{L,s}}{\dfrac{3}{\pi} \dfrac{E_m}{R_0}} = 1 - \sin\left(-\frac{\dfrac{I_C N_c}{AT_c} + \bar{h}_c'}{\dfrac{3}{\pi}(1 - \bar{h}_c')} + \frac{\pi}{6}\right) \quad (72)$$

The maximum load current $I_{L,m}$ with fully saturated reactors is

$$I_{L,m} = \frac{3}{\pi} \frac{E_m}{R_0} \quad (73)$$

as will be proved by equation 84 in section *D*. Substituting from equations 72 and 73 one obtains in nondimensional form the control characteristic concerning the saturating components, as shown in Fig. 12(B) by the cosine, from $I_C N_C / AT_c = -1$ to N.

To obtain the total control characteristic the two characteristics in Fig. 12(A) and (B) must be added, after the application of a proper scale factor. Let the per-unit exciting current $\imath_x$ be defined as

$$\imath_x = \frac{2AT_c}{I_{L,m} N_G} \quad (74)$$

Then, the ordinates of Fig. 12(A), when multiplied by $\imath_x$, can be added to the ordinates of Fig. 12(B). In practical cases the per-unit excitation $\imath_x$ is in the vicinity of 1% and therefore the effect of the exciting components is of practical significance only at or near minimum load. Actually, measured points are also shown in Fig. 12(B) by crosses. In the range from minimum load to half load (*MN*) the control circuit resistance can be varied from $R_C = 600$ to 26,000 ohms without noticeable effects.

A fundamental quantity for the evaluation of amplifier performance is the ampere-turn gain K_{AT}

$$K_{AT} = \frac{d(I_L N_G)}{d(I_C N_C)} \quad (75)$$

Since the contribution of the exciting component of the load ampere-turns (equation 71) is so small, only the saturation component will be considered. By differentiating equation 72 with respect to $I_C N_C$, one obtains

$$K_{AT} = \frac{I_{L,m} N_G}{2AT_c} \frac{2\pi}{3} \frac{1}{1 - \bar{h}_c'} \cos(\alpha_s - 30^\circ)$$
$$60^\circ \leq \alpha_s \leq 120^\circ \quad (76A)$$

The highest value for the ampere-turn gain $K_{AT,m}$ is obtained for $\alpha_s = 60^\circ$, resulting in

$$K_{AT,m} = \frac{I_{L,m} N_G}{2AT_c} \frac{\pi}{\sqrt{3}} \frac{1}{1 - \bar{h}_c'} \qquad \alpha_s = 60^\circ \quad (76B)$$

In Figs. 13 through 17 the saturation angle $\alpha_s = 78.5^\circ$. With this angle, $I_{L,s}$ becomes 25% (equations 68 and 73) of $I_{L,m}$. For comparison purposes oscillograms are shown in Table I, for 25% load current. Oscillograms are also shown for the boundary case of 50% load current.

D. Half Load to Full Load, $0^\circ \leq \alpha_s \leq 60^\circ$ Normal Excitation

"Half load to full load" refers to the magnitude of the load current which will be varied in this section from one half to the full value of $I_{L,m}$ (*NP* in Fig. 12(B)). This variation of load current is achieved with constant-load resistance R_L, by varying the saturation angle α_s from 60° to 0°; see equation 86.

It was shown in section C1 that raising the starting flux Φ_1 from $-\Phi_s$ to 0 changes the saturation angle α_s from 120° to 60°, with an attendant decrease in the duration of mode 1 and an increase in the duration of mode 2. For $\alpha_s = 60^\circ$, only mode 2 exists. If the starting flux is raised to become positive, the saturation angle becomes smaller than 60°, and two modes of operation will be encountered: mode 2 for $0^\circ \leq \alpha \leq \alpha_s$ (identical with mode 2 of section C), and a new mode, called mode 3 for $\alpha_s \leq \alpha \leq 60^\circ$.

1. Reactor and Rectifier Voltages

Mode 2 $(0^\circ \leq \alpha \leq \alpha_s)$

The voltage and current configurations are shown for $\alpha_s = 41^\circ$ (75% loading) in Figs. 18(A) and (B): $300^\circ \leq \theta \leq 360^\circ$. Core A is setting and is resetting cores C', B, and A' by equal amounts similar to those shown in section C1 and equation 53(B). The remaining cores C and B' are saturated and therefore produce no voltages.

Mode 3 $(\alpha_s \leq \alpha \leq 60^\circ)$

It will be seen from Fig. 16 (minimum to half load) that cores C and B' were saturated for $(\theta_s - 60^\circ) \leq \theta \leq 0^\circ$. As the saturation angle α_s is reduced below 60°, core A also becomes saturated, which means that there are now three cores which are saturated. The other three cores are idle, as shown in Table II.

Entire Interval

The gate voltages can now be determined for the entire period; see Fig. 19(A).

The rectifier reverse voltages are determined in the same manner as in section C1, and are shown in Fig. 19(B). It will be seen that the maximum rectifier reverse voltage equals the crest value of the line-to-line voltage:

$$E_{r,m} = E_m \quad (77)$$

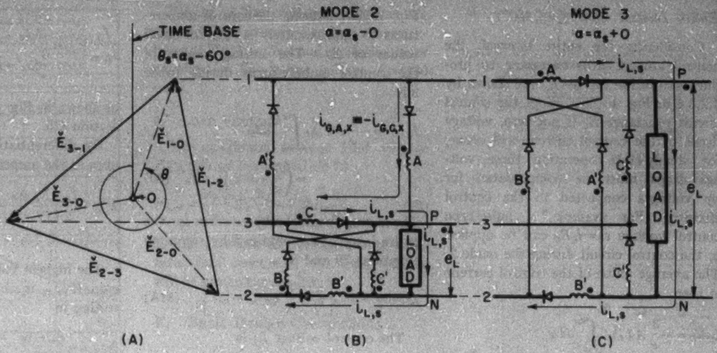
Vol.
Pt.
195

Fig. 18. Bus voltages at 75% load ($\alpha_s=41°$)

A—Phasor diagram. Accented E's indicate phasors

B—Bus voltages and currents just before saturation

C—Bus voltages and currents just after saturation



2. Fluxes

The core flux of A is given again by equation 58(A), as shown in section C2. For the remaining part of the period, core flux ϕ_A is determined by a step-by-step integration of equation 30 with the gate voltage $e_{G,A}$, taken from Fig. 19(A). Core flux ϕ_A is shown for the entire period in Fig. 19(C).

3. Exciting Components of Currents Including Control Current

Mode 2

The gate current $i_{G,A,x}$ in this subinterval $(300° \leq \theta \leq 300°+\alpha_s)$ has the same make-up as gate current $i_{G,C',x}$ had in section C3, $(\theta_s \leq \theta \leq 60$, Fig. 17(A), equation 60(B):

$$i_{G,A,x}N_G=AT_c(1+\bar{h}_c') \tag{78}$$

This current can return as $-i_{G,C,x}$ to bus 3 via the reverse direction of rectifier C because this rectifier is kept open in the forward direction by the flow of load current from bus 3 to bus 2. During $60° \leq \theta \leq 60+\alpha_s$, current $i_{G,C,x}$ becomes identical to $-i_{G,A,x}$, resulting in exciting components of the gate currents, shown in Fig. 19(D). Fig. 18(B) also shows that the exciting components avoid the load circuit and therefore $i_{L,x}=0$; see Fig. 19(F).

The control ampere-turns follow from equation 59 and can be seen in Fig. 19(H).

Mode 3

No cores are being excited, therefore $i_{G,A,x}=0$ Fig. 19(D), and $i_{L,x}=0$; see Fig. 19(F). If no voltages are induced into the control circuit from the gates or applied by a control voltage source, the control ampere-turns area 4,5,6,7, Fig. 19(H), will also be zero.

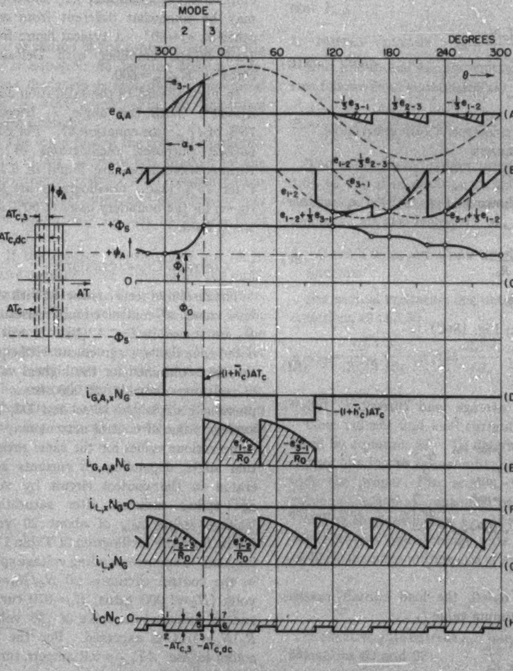


Fig. 19. Saturation angle $\alpha_s=41°$ (75% load)

A—Gate voltage

B—Rectifier voltage

C—Core flux

D—Exciting component of gate current

E—Saturation component of gate current

F—Exciting component of load current

G—Saturation component of load current

H—Control current

Entire Interval ($300° \leq \theta \leq 360°$)

Considering the entire interval, the control ampere-turns necessary to produce saturation for $\alpha = \alpha_s$, are shown by 1,2,3,4 in Fig. 19(H). Since the control circuit resistance R_C is not zero, voltage drops of the control current will occur; for steady-state operation these voltage drops must be compensated for by voltages connected to the control circuitry. For example, a pulse-type control voltage $e_C = i_C R_C$ can be applied to the control circuit during the mode 2. The average value of the control current is then

$$I_C N_C = -\frac{3}{\pi} AT_c \bar{h}_c' \int_0^{\alpha_s} \alpha d\alpha$$
$$= -\frac{3}{\pi} \alpha_s AT_c \bar{h}_c' \quad (79)$$

The control voltage could be modified by attaching to it voltages existing during mode 3. If these added-on voltages are kept below $AT_c\ (R_C/N_C)$, the resulting control currents during mode 3 will not change the saturation angle or power output, yet the average value of $I_C N_C$ (equation 79) may have been changed. One concludes that the output of the amplifier is determined jointly by the average value $I_C N_C$ of control ampere-turns, and by the magnitude and shape of the applied control voltage.

In this paper it has been assumed that no coupling exists between gate and control windings, once the reactor saturates. Actually this is not quite so, as is borne out by the various voltage spikes seen in the oscillograms of Table I. These voltages spikes produce currents, which, if large enough, may operate on the **B-H** loop, and thereby change the average value of control ampere-turns $I_C N_C$ for a given saturation angle even if the voltage spikes are pure alternating current. These voltages spikes become larger for a decreasing saturation angle α_s, because of the concomitant increase of gate current. It will also be seen from Fig. 19(A) that a decreasing saturation angle reduces the duration and the magnitude of the regular, sinusoidal reset voltage derived from the line-to-line voltage. Hence, with decreasing saturation angle α_s, the voltage spikes exert increasing relative influence, and thereby cause deviations from the control characteristic of Fig. 12(B).

In the following, the control characteristic will be calculated under conditions of a constant control voltage E_C, and no induced voltages after saturation.

During mode 3, the constant control voltage produces a control current 4,5,6,7, Fig. 19(H), limiting the control ampere-turns to approximately $AT_{c,dc}$ for small values of E_C. The average value of the control ampere-turns during mode 3 is

$$I_C N_C = -\frac{3}{\pi} AT_c h_c \int_{\alpha_s}^{\pi/3} \alpha d\alpha$$
$$= \left(\frac{3}{\pi} \alpha_s - 1\right) \bar{h}_c AT_c \quad (80)$$

The average control ampere-turns for the entire interval are obtained by summing equations 79 and 80:

$$I_C N_C = -AT_c \left[\bar{h}_c + \frac{3}{\pi} \alpha_s (\bar{h}_c' - \bar{h}_c)\right] \quad (81A)$$

The control voltage E_C is

$$E_C = I_C R_C \quad (81B)$$

and the saturation angle α_s from equation 81(A)

$$\alpha_s = -\frac{\dfrac{I_C N_C}{AT_c} + \bar{h}_c}{\dfrac{3}{\pi}(\bar{h}_c' - \bar{h}_c)} \qquad 0° \leq \alpha_s \leq 60° \quad (81C)$$

4. SATURATION COMPONENTS OF CURRENTS

Mode 2

As shown in Fig. 18(B), the load current $i_{L,s}$ is

$$i_L = -\frac{e_{2-3}}{R_0} \quad (82)$$

Mode 3

From Fig. 18(C)

$$i_L = \frac{e_{1-2}}{R_0} \quad (83)$$

The average load current I_L follows from

$$I_L = \frac{3}{\pi}\left[\int_{300°}^{300°+\alpha_s} -\frac{e_{2-3}}{R_0} d\theta + \int_{300°+\alpha_s}^{360°} \frac{e_{1-2}}{R_0} d\theta\right] = \frac{3}{\pi} \frac{E_m}{R_0} \cos \alpha_s \quad (84)$$

As there is no exciting component, $I_{L,s} = I_L$.

For $\alpha_s = 0$, the load current reaches its maximum value $I_{L,m}$:

$$I_{L,m} = \frac{3}{\pi} \frac{E_m}{R_0} \quad (85)$$

In nondimensional form, the load current is expressed by

$$\frac{I_L}{I_{L,m}} = \cos \alpha_s \quad (86)$$

In terms of core properties, one obtains the control characteristic by substituting from equation 81(C) into 86:

$$\frac{I_L}{I_{L,m}} = \cos \frac{\dfrac{I_C N_C}{AT_c} - \bar{h}_c}{\dfrac{3}{\pi}(\bar{h}_c' - \bar{h}_c)} \quad (87)$$

as shown in Fig. 12(B), and discussed in section C5.

By differentiation of equation 87 one obtains the ampere-turn gain:

$$K_{AT} = \frac{I_{L,m} N_G}{2AT_c} \frac{2\pi}{3} \frac{1}{(\bar{h}_c' - \bar{h}_c)} \sin \alpha_s$$
$$0° \leq \alpha_s \angle 60° \quad (88A)$$

The highest value for the ampere-turn gain $K_{AT,m}$ is obtained for $\alpha_s = 60°$, resulting in

$$K_{AT,m} = \frac{I_{L,m} N_G}{2AT_c} \frac{\pi}{\sqrt{3}} \frac{1}{\bar{h}_c' - \bar{h}_c} \quad (88B)$$

It will be noticed that the maximum ampere-turns, equations 76(B) and 88(B), may be somewhat different from each other at $\alpha_s = 60°$. A typical figure for a 1-kw 60-cps amplifier with Deltamax cores is $K_{AT,m} = 600$.

In Figs. 18 and 19 the saturation angle $\alpha_s = 41°$. With this angle, I_L becomes 75% of $I_{L,m}$; see equation 87. For comparison purposes, oscillograms of the various waveshapes are found in Table I for 75% load. Oscillograms are also shown for the boundary cases of 50% and 100% load.

5. CONTROL CHARACTERISTIC

The results of tests conducted with various values of control circuit resistances R_C are shown in Fig. 12(B). It will be noted that the best agreement with equation 86 is obtained for the highest value of resistance, namely, 26,000 ohms. As previously explained in section D3, the average value of control current may assume various values for the same saturation angle, depending on currents generated in the control circuit by voltage spikes induced after saturation. Voltage spikes $e_{G,p}$ of about 20 volts follow from the oscillogram of Table I for full load. The corresponding voltage spike in the control circuit is 20 $N_C/N_G = 30$ volts ($N_C = 1{,}000$ turns, $N_G = 670$ turns) and an ampere-turn spike of (30 volts/R_C)$N_C = 1.14$ is expected. For the reactors in use, $AT_{c,dc} = 2.8$ ampere turns, and therefore the 1.14 ampere-turn spike cannot produce a significant flux change. Hence, the average control ampere-turns as shown in Fig. 12(B) will be hardly affected. By reducing R_C to 12,000 ohms, the induced spikes become 2.48 ampere-turns, still being less than $AT_{c,dc}$. However, for $R_C = 6{,}000$ ohms, the spikes become 4.96 ampere turns, and

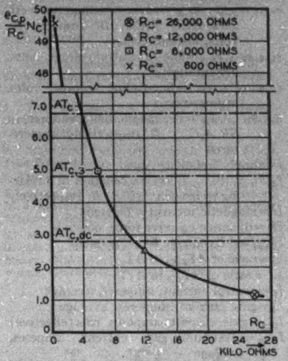


Fig. 20. Magnitudes of ampere-turn spikes $e_{C,P} N_C/R_C$ for various values of control circuit resistance R_C

the deviation of the control characteristic from equation 86 become considerable. This deviation is still more pronounced for $R_C=600$ ohms. Fig. 20 illustrates these relations.

The lower half of the control characteristic in Fig. 12(B) is not affected by the variation of control circuit resistance from 600 to 26,000 ohms, and measured points are shown only for one value ($R_C=600$ ohms).

If the control ampere-turns are made more negative than required for the minimum load point, the control characteristics of Fig. 21 result. The slopes of the characteristics are negative and the ampere-turn gain very small ($K_{AT}=2$ for the lower part of the characteristic). The control characteristic of Fig. 12, when drawn to the scale of Fig. 21 becomes the line coincident with the ordinate axis.

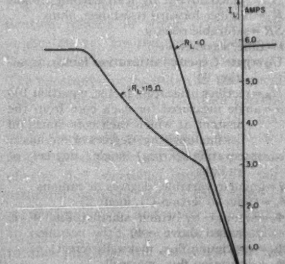


Fig. 21. Control characteristics of amplifier for the negative region

Effects of line voltage variations are shown by the measured control characteristics of Fig. 10. Normal excitation is for an effective line-to-line voltage $E_s=150$ volts.

6. LOAD VOLTAGE

The maximum average load voltage $E_{L,m}$ follows from equation 73:

$$E_{L,m}=\frac{3\sqrt{2}}{\pi}E_s\frac{R_L}{R_0}=1.35E_s\frac{R_L}{R_0} \tag{89}$$

In the presence of overlap due to post-saturation reactance, the load voltage will be somewhat smaller.

E. Basic Design Calculations

Usually the output requirement of an amplifier is given by the maximum load voltage $E_{L,m}$ and the maximum load current $I_{L,m}$.

1. SUPPLY VOLTAGE

The line-to-line supply voltage E_s in rms follows from equation 89:

$$E_s=\frac{\pi}{3\sqrt{2}}E_{L,m}\frac{R_0}{R_L}=0.74E_{L,m}\frac{R_0}{R_L} \tag{90}$$

The ratio R_0/R_L ranges usually from 1.01 to 1.3, if in addition to the circuit resistances, the effects of post-saturation reactance are also included.

2. GATE TURNS AND CORE CROSS SECTION

For normal excitation one obtains from equations 49 and 50

$$A_{Fe}N_G=\frac{10^8}{2\sqrt{2}}\frac{E_s}{fB_s}=0.1125\frac{E_s10^8}{fB_s} \tag{91}$$

3. GATE CURRENT

Gate current and load current are related by equation 69. To determine the copper loss in the gate winding, and provide the proper cross section, the rms value of the gate current must be known. Defining the form factor $k_{f,G}$ as the ratio between rms and average gate current yields

$$k_{f,G}=\frac{I_{G,e}}{I_G} \tag{92}$$

The effective gate current $I_{G,e}$ follows from equations 69 and 92:

$$I_{G,e}=k_{fG}I_G=\frac{1}{3}k_{f,G}I_L \tag{93}$$

Measured values for $k_{f,g}$ are shown in Fig. 22. At full load, the gate current resembles a rectangle; the form factor of a reactangle of one third of a period duration is $\sqrt{3}$.

From equations 91 and 93, the core size and the proper winding can be determined, as shown, for instance, by Anderson.[16,17]

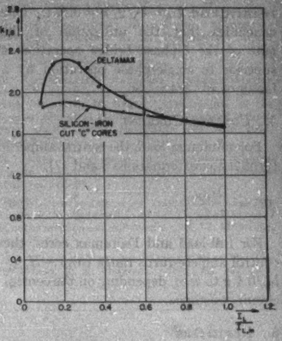


Fig. 22. Form factor for gate currents

4. UTILIZATION OF REACTOR

The reactor utilization may be defined as the share of output of this reactor divided by the volt-amperes of the reactor gate. For the circuit of Fig. 1, this ratio U_G becomes for full load and zero losses

$$U_G=\frac{\frac{1}{6}E_{L,m}I_{L,m}}{\frac{1}{2}E_sI_{G,e}} \tag{94}$$

Substituting for E_s from equation 90 and for $I_{G,e}$ from equation 93, one obtains with $k_{f,G}=\sqrt{3}$

$$U_G=\frac{\sqrt{6}}{\pi}=0.78 \tag{95}$$

For the single-phase amplifier with

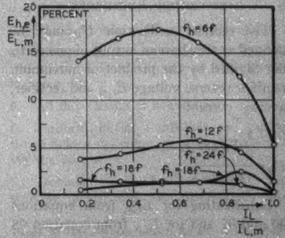


Fig. 23. Measured load voltage ripple of amplifier with Deltamax cores

$E_{h,e}$ = rms value of harmonic voltage
$E_{L,m}$ = rated d-c load voltage
f = supply frequency
fh = harmonic frequency

resistive load, $U_G=4\sqrt{2}/\pi^2=0.574$. One concludes that the utilization of the gate of the 3-phase bridge operation is superior to single-phase operation by a factor of $\pi\sqrt{3}/4=1.36$.

5. CONTROL CURRENT

For minimum load, the control ampere-turns AT_C are (equations 8 and 17)

$$AT_c=-\frac{H_c l_{Fe}}{0.4\pi} \tag{96}$$

For full-load and Deltamax cores, the control ampere-turns range from $-H_{c,dc}l_{Fe}/0.4\pi$ to zero, depending on the control circuit resistance.

6. RECTIFIERS

The maximum rectifier reverse voltage $E_{r,m}$ follows from equations 48 and 77:

$$E_{r,m}=\sqrt{2}E_e \tag{97}$$

The maximum average rectifier current $I_{G,m}$ follows from equation 69:

$$I_{G,m}=\frac{1}{3}I_{L,m} \tag{98}$$

In contrast to mercury arc or gaseous rectifiers, the rating of a semiconductor rectifier is determined not only by the average value, but also by the circuit connection as evidenced in manufacturers' catalogs.

7. LINE CURRENT

The line current i_l consists of one positive current train such that as shown in Fig. 19(E), followed 180° later by the same waveshape, but with reverse polarity. For full load, these currents can be approximated by rectangles, resulting in the simple relation

$$I_{l,e}=\sqrt{\frac{2}{3}}I_{L,m}=0.816I_{L,m} \tag{99}$$

8. UTILIZATION OF RECTIFIERS

The rectifier utilization U_R may be defined as the share of output of one rectifier divided by the product of maximum, rectifier reverse voltage $E_{r,m}$ and rectifier, average current

$$U_R=\frac{\frac{1}{6}E_{L,m}I_{L,m}}{E_{r,m}I_{G,m}} \tag{100}$$

Substituting for $E_{r,m}$ from equations 90 and 97 and for $I_{G,m}$ from equation 98 yields

$$U_R=\frac{3}{2\pi}=0.477 \tag{101}$$

For single-phase operation, $U_R=1/\pi=0.318$. Hence, the rectifier appears to be utilized better in a 3-phase bridge connection by a factor of 3 to 2. Actually, the average current rating of a semiconductor rectifier in a 3-phase bridge connection is about 10% less than in single-phase connection; therefore the superiority of the 3-phase bridge connection becomes $0.9\times3/2=1.35$

9. POWER FACTOR

The power factor (pf) relates the maximum d-c output power $E_{L,m}I_{L,m}$ (assuming no losses) to $E_eI_{l,e}\sqrt{3}$

$$pf=\frac{E_{L,m}I_{L,m}}{E_eI_{l,e}\sqrt{3}} \tag{102}$$

For full load and assuming rectangular gate currents, one obtains from equations 90 and 99

$$pf=\frac{3}{\pi}=0.955 \tag{103}$$

10. LOAD RIPPLE

To enable the designer to calculate load filters, the ripple of the load voltage was measured. Deltamax cores were used; the results are shown in Fig. 23 for the 6th, 12th, 18th, and 24th harmonics. The minimum ripple of the single-phase amplifier is 48%.

The comparisons listed above are intended to point out some of the inherent advantages of the 3-phase bridge amplifier over the single-phase amplifier.

Nomenclature

A_{Fe} = area of active core iron, square centimeters
AT = ampere-turns
AT_C = control ampere-turns
AT_c = ampere-turns to produce coercive force H_c
$AT_{c,dc}$ = ampere-turns to produce static hysteresis loop
$AT_{c,s}$ = (for definition see section C3)
B = flux density, gausses
B_s = saturation flux density, gausses
E_C = control voltage, volts, average
$E_{G,A}, E_{G,C'}, \ldots E_{G,B'}$ = gate voltages on SR $A,C'\ldots B'$ respectively, volts, average
$E_{L,m}$ = maximum load voltage, volts, average
E_e = line-to-line voltage, volts, rms (effective)
E_m = line-to-line voltage, volts, maximum (crest)
$E_{r,m}$ = rectifier reverse voltage, volts, maximum (crest)
$e_{C,A}, e_{C,C'}\ldots e_{C,B'}$ = voltage across control winding of SF $A,C'\ldots B'$ respectively, volts, instantaneous
$e_{C,p}$ = voltage spike induced into control circuit, volts; see (section D.5 and Fig. 20
$e_{G,A}, e_{G,C'}, \ldots e_{G,B'}$ = voltage across gate winding of SR $A,C'\ldots B'$ respectively, volts, instantaneous
$e_{R,A}, e_{R,C'}, \ldots e_{R,B'}$ = rectifier voltage of REC $A,C',\ldots B'$ respectively, volts, instantaneous
$e_{r,A}, e_{r,C'}, \ldots e_{r,B'}$ = reverse voltage on rectifiers REC $A,C'\ldots B'$ respectively, volts, instantaneous
$e_{1-0}, e_{2-0}, e_{3-0}$ = line-to-neutral voltage, volts, instantaneous
$e_{1-2}, e_{2-3}, e_{3-1}$ = line-to-line voltage, volts, instantaneous
$\mathfrak{F}_A, \mathfrak{F}_C'\ldots \mathfrak{F}_B'$ = net magnetomotive force on SR $A,C',\ldots B'$ respectively, ampere-turns
f = supply frequency, cps
$f_A(\theta), f_C'(\theta), \ldots f_B'(\theta)$ = a function of θ relating to branch $A,C'\ldots B'$ respectively
H = magnetic intensity, oersteds
H_c = dynamic coercive force, oersteds
$H_{c,dc}$ = d-c coercive force, oersteds
$\bar{h}_c$ = ratio of $AT_{c,dc}$ to AT_c
$\bar{h}_c'$ = ratio of $AT_{c,s}$ to AT_c
I_C = control current, amperes, average
I_G = gate current, amperes, average
$I_{G,e}$ = gate current, amperes, rms (effective)
$I_{G,m}$ = maximum gate current, amperes, average
$I_{G,s}$ = saturation component of gate current, amperes, average
I_L = load current, amperes, average
$I_{L,m}$ = maximum load current, amperes, average
$I_{L,s}$ = saturation component of load current, amperes, average
$I_{L,x}$ = excitation component of load current, amperes, average
$I_{l,e}$ = line current, amperes, rms (effective)
i_C = control current, amperes, instantaneous
$i_{G,A}, i_{G,C'}, \ldots i_{G,B'}$ = gate current, amperes, instantaneous
$i_{G,A,x}, i_{G,C',x}, \ldots i_{G,B',x}$ = excitation component of gate current, amperes, instantaneous
i_L = load current, amperes, instantaneous
$i_{L,x}$ = excitation component of load current, amperes, instantaneous
$(iN)_x$ = exciting ampere-turns
$\bar{i}_x$ = per-unit exciting current
K_{AT} = ampere-turn gain
$K_{AT,m}$ = ampere-turn gain maximum value
$k_{f,G}$ = form factor relating the rms to average gate current
l_{Fe} = length of magnetic circuit, centimeters
N = number of turns
N_C = number of turns of one control winding
N_G = number of turns of one gate winding
pf = power factor of load
R_C = resistance of entire control circuit, ohms
R_G = resistance of gate winding, ohms
R_L = load resistance, ohms
R_0 = output resistance (total resistance encountered by load current), ohms
R_f = rectifier forward resistance, ohms
SR = saturable reactor
t = time, seconds
U_G = gate (reactor) utilization factor, equation 94
U_R = rectifier utilization factor, equation 100
α = angle measured for each core from the instant at which each core starts on its flux upswing, degrees or radians
α_s = saturation (firing) angle, degrees or radians
θ = angular variable, degrees or radians
θ_u = (refer to section C1, mode 2)
Φ_0 = amount by which starting flux Φ_1 is raised above $-\Phi_s$
Φ_m = maximum flux, maxwells (crest)
Φ_s = saturation flux, maxwells
Φ_1 = starting flux, maxwells
$\phi_A, \phi_C', \ldots \phi_B'$ = core flux in SR $A, C'\ldots B'$ respectively, maxwells, instantaneous
$\Delta\phi$ = incremental change of core flux, maxwell, instantaneous

ω = angular velocity of line veltage, radians per second

References

1. The Transductor Amplifier (book), **U. H. Krabbe.** Lindhska Boktrycheriet, Orebro, Sweden, 1947.

2. The Transductor (book), **A. U. Lamm.** Esselte Aktiebolaget, Stockholm, Sweden, second edition, 1948.

3. Magnetic Amplifiers (book), **H. F. Storm.** John Wiley & Sons, Inc., New York, N. Y., 1955, p. 359.

4. Magnetic Amplifiers (book, in Russian), **M. A. Rosenblat.** Isdatelstvo "Sovietskoe Radno," Moscow, USSR, 1956, p. 577.

5. Magnetic Amplifiers (book, in Czechoslovakian), **M. Gabler, J. Haškovec, E. Tomášek.** Státní Nakladatelstvi Technické Literatury, Prague, Czechoslovakia, 1956, p. 359.

6. Transductors and Magnetic Amplifiers (book), **A. G. Milnes.** Macmillian and Company, Ltd., London, England, 1957, pp. 181–207.

7. Dynamic Behavior of a 3-Phase Magnetic Amplifier, **F. J. Ellert.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, Nov. 1956, pp. 554–60.

8. The Polyunit Saturable Reactor, **Karl I. Selin.** *Ibid.*, pt. III (*Power Apparatus and Systems*), vol. 75, Oct. 1956, pp. 863–67.

9. Experimental Characteristics of the 3-Phase Polyunit Saturable Reactor, **Karl I. Selin, Alexander Kusko.** *Ibid.*, pp. 868–71.

10. Magnetic Amplifiers for High-Performance Power Servos, **G. P. DeWestfelt.** *Electrical Manufacturing*, New York, N. Y., Nov. 1956, p. 114.

11. The Power Magnestat, **S. Rudzinski.** *B.T.H. Activities*, Rugby, England, vol. 27, no. 5, Sept.-Oct. 1956, pp. 154–58.

12. Three-Phase High-Speed Magnetic Amplifiers, **A. E. Maine.** *Electronic Engineering*, London, England, Dec. 1954, pp. 514–21.

13. 160,000-Ampere High-Speed Magnetic-Amplifier Design, **A. B. Rosenstein.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, Mar. 1955, pp. 90–99.

14. Generalized Transfer Characteristics for 3-Phase Magnetic Amplifiers with Inductive Loads, **C. H. Grace.** "Proceedings of the Special Technical Conference on Magnetic Amplifiers," *AIEE Special Publication T98*, Aug. 1957, pp. 169–76.

15. New Techniques in Designing Three-Phase Magnetic Amplifiers, **W. H. Moore.** *Ibid.*, pp. 177–86.

16. Magnetic-Amplifier Design, Part I—The Core, **R. E. Anderson.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, May 1958, pp. 160–67.

17. Magnetic Amplifier Design, Part II—The Windings, **R. E. Anderson.** *Ibid.*, pp. 167–75.

Two-Step Switching of Saturable Reactors

B. W. LOVELL
ASSOCIATE MEMBER AIEE

THE FUNDAMENTAL problem in the analysis and design of magnetic amplifiers is the determination of the unsaturated impedance of the reactor by some test less specialized than insertion in the actual circuit. For tape and ferrite cores driven by very large currents, such as are encountered in computer applications, fairly successful approaches have been developed, although they are still incapable of explaining some observed phenomena.[1–3] Unfortunately, cores operating at normal magnetic-amplifier frequencies are not correctly described by these representations, and the designer has been forced to use either extremely idealized models or fairly limited extrapolations from specific measurements.

A core model based upon the magnetization process in single crystals, and applicable to low-frequency operation, has recently been reported by Woodson.[4] The constants incorporated in the model were evaluated from constant-current switching characteristics. For the material Orthonol* this model was shown to give good predictions of the operation of the d-c controlled single-core self-saturating amplifier over a wide range of carrier frequencies and input impedances. In common with most other approaches to this problem, the excitation for the model was assumed to be only slowly varying.[5,6]

This paper investigates the ability of that model and of its simplified forms to handle the more violently fluctuating excitations characteristic of multicore and polyphase magnetic-amplifier circuits. The particular situation considered is flux reversal by a step of magnetic-field intensity which abruptly changes amplitude during the reversal process. The applied fields are as shown in Fig. 1.

Mathematical Model

The behavior of an element of domain wall in a single crystal has been found to be described by equations of the form:

$$f(\delta)\frac{d\delta}{dt}=K_1(H-H_s) \qquad (1)$$

$$\frac{d\phi_1}{dt}=2B_s\frac{dA_1}{d\delta}\frac{d\delta}{dt} \qquad (2)$$

where

δ = distance of domain-wall element from nucleation site
$F(\delta)$ = position dependent damping factor
K_1 = material constant
H = applied field intensity
H_s = starting field for the element of domain wall
$\frac{d\phi_1}{dt}$ = rate of change of flux produced by this domain-wall element
B_s = saturation flux density
$\frac{dA_1}{d\delta}$ = relationship between area perpendicular to domain wall which is swept out by the element of wall and the position of the element

These equations were averaged over all domain walls by Woodson to produce two equations describing the switching process of polycrystalline core:[4]

$$\frac{dr}{dt}=K_1K_2\int_{H_c}^{H}(H-H_s)G(H_s)dH_s \qquad (3)$$

$$\frac{d\phi}{dt}=2B_sA_cg(r)\frac{dr}{dt} \qquad (4)$$

where

r = average normalized domain dimension
K_2 = normalizing constant
$G(H_s)$ = number of elements of domain wall having starting fields between H_s and H_s+dH_s
H_c = coercive force of core, the lowest value of H_s
$\frac{d\phi}{dt}$ = total rate of change of flux
A_c = effective cross-sectional area of core
$g(r)$ = shape function relating rate of change of flux to instantaneous r

In the development, the following assumptions were used or implied:

1. The magnetic field intensity is uniform throughout the material.
2. The nucleating field for each domain is less than its starting field.
3. The d-c hysteresis loop is rectangular.
4. All domains have starting fields $H_s \geq H_c$ which are constant throughout the switching process.
5. All flux change is irreversible.
6. Uniform initial conditions are established before switching by strongly saturating the core.
7. Domains are identical except for starting fields and are distributed uni-

---

* Trade-mark of Magnetics, Inc., Butler, Pa., for grain-oriented annealed 50 nickel-50 iron alloy.

Paper 59-1050, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-IRE Nonlinear Magnetics and Magnetic Amplifier Conference, Washington, D. C., September 23–25, 1959. Manuscript submitted November 12, 1958; made available for printing July 16, 1959.

B. W. Lovell is with Massachusetts Institute of Technology, Cambridge, Mass.

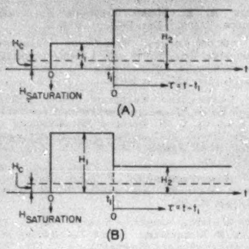


Fig. 1. Applied fields in 2-step switching process

A—Increase in applied field
B—Decrease in applied field

formly. The distribution of starting fields is independent of the applied field.

8. Each domain is described by equations of the form of 1 and 2.

9. $F(\delta)=1$.

10. The rate of change of applied field is small.

11. The rate of change of flux depends only upon the value of r and upon its rate of change.

For constant field switching of Orthonol, it was stated that $g(r)$ could be approximated by a sinusoid,

$$g(r)=\frac{\pi}{2}\sin \pi r \tag{5}$$

and that over a wide range of switching speeds $G(H_s)$ was a constant, a_0.[4] With these simplifications, the describing equation for the core becomes:

$$\frac{d\phi}{dt}=\pi B_s A_c K_1 K_2 \frac{a_0}{2}(H-H_c)^2\times \sin\left[\pi K_1 K_2 \frac{a_0}{2}(H-H_c)^2 t\right] \tag{6}$$

The model was extended to the case of slowly varying applied field by using equation 6 on an instantaneous basis in accordance with assumption 11.[4] This type of assumption is clearly not valid for all cases of discontinuous excitation. The growth of new domains after an increase in excitation cannot be as extensive as it would have been had the large excitation been applied initially, because some of the potential nucleating sites will be included in the region where magnetization has already been reversed at the time of the increase. Because domains of all starting fields are assumed uniformly distributed, the fractional decrease in the number of new domains is that fraction of the core where magnetization has been reversed by the time of the increase. This modification gives a core memory of how it reached a given flux level.

If domains grow from very small nuclei with a velocity dependent only upon the difference between the applied field and the starting field, as assumed, then newly formed domains should not contribute a large switching voltage. To bring the model into line with this consideration while adhering as closely as possible to the original analysis, the sinusoidal shape function which provides initial growth and final collision patterns will be applied to the individual domains. For switching processes during which the excitation is increased, all domains will be considered to grow from the time when their starting field is reached until the core is saturated.

No new domains are formed by a decrease in excitation during the switching process, and the critical assumption is that domain-wall velocity is determined only by the applied field and the starting field of the wall, regardless of the size that the domain may have attained.

The use of the sinusoidal shape function with equation 1 and equation 2 reduces to:

$$\frac{d\phi_1}{dt}=2B_s K_1(H-H_s)K_3 \sin\left(\pi\frac{\bar{\delta}}{\bar{\delta}^*}\right) \tag{7}$$

where

$\bar{\delta}$ = the average distance of all domain-wall elements from their nucleation sites
$\bar{\delta}^*$ = the final value that $\bar{\delta}$ is expected to attain because of present and past excitation
K_3 = a constant to be evaluated so that peak constant-current switching voltage is correct

After the applied field has been changed from one constant value to another, this becomes:

$$\frac{d\phi_1}{dt}=2B_s K_1(H-H_s)K_3\times \sin\left[\pi\left(\frac{\bar{\delta}_1}{\bar{\delta}_1^*}+\frac{\bar{\epsilon}}{\bar{\delta}_2^*}\right)\right](H>H_s) \tag{8}$$

where

$\bar{\delta}_1^*$ = the final value $\bar{\delta}$ was expected to reach before the change in excitation
$\bar{\delta}_1$ = the value $\bar{\delta}$ had reached at the time of the change
$\bar{\delta}_2^*$ = the final value $\bar{\delta}$ does reach at saturation
$\bar{\epsilon}$ = the amount by which $\bar{\delta}$ has changed since the excitation was altered

By equation 1, $\bar{\delta}$ is increasing linearly with time both before and after the change, but with a different coefficient. That is,

$$\frac{\bar{\delta}_1}{\bar{\delta}_1^*}=\frac{t_1}{t_1^*} \tag{9}$$

$$\frac{\bar{\epsilon}}{\bar{\delta}_2^*}=\left(1-\frac{t_1}{t_1^*}\right)\frac{\tau}{t_2^*} \tag{10}$$

where

t_1 = time at which the change in excitation occurs. Time is measured from the start of the flux reversal process
t_1^* = the time when the core would have saturated had the excitation not been changed
t_2^* = the time between the increase in excitation and saturation
τ = time since the change in excitation

Equation 8 applies to already growing domains for both increases and decreases in excitation. The same result may be obtained by replacing the average values in equation 8 with the values pertaining to the particular domain considered and using assumption 11 as applied to that domain only; i.e., utilizing the idea that any two of the quantities

$$\phi_1,\ \frac{d\phi_1}{dt},\ \text{and}\ (H-H_s)$$

determine the other. For domains which start to grow after an increase in excitation, the shape function is taken to be $K_3 \sin(\pi\,\tau/t_2^*)$ so that the initial and final growth patterns correspond to those of the model. Finally, integrating over all domains, the describing equations for a first level of excitation H_1, and a second level H_2, are

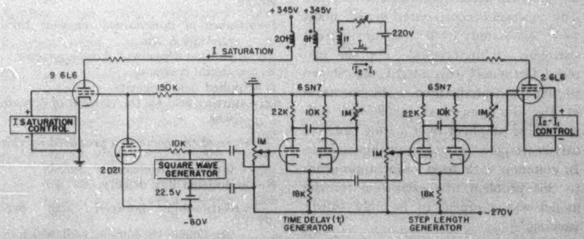


Fig. 2. Excitation circuit

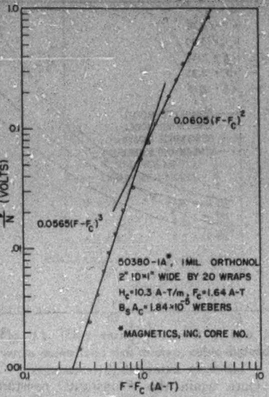


Fig. 3. Peak constant field switching voltage versus excitation

$$\frac{d\phi}{dt}=2B_sK_1K_2\int_{H_c}^{H_1}(H_1-H_s)G(H_s)\times \sin\left(\pi\frac{t}{t_1^*}\right)dH_s \quad (0<t<t_1) \qquad (11)$$

$$\frac{d\phi}{dt}=2B_sK_1K_2\int_{H_c}^{H_1}(H_1-H_s)G(H_s)\times \sin\left[\pi\left(\frac{t_1}{t_1^*}+\frac{t_1^*-t_1}{t_1^*\ t_2^*}\tau\right)\right]dH_s+ \frac{1}{2}\left(1+\cos\pi\frac{t_1}{t_1^*}\right)\int_{H_1}^{H_2}(H_2-H_1)\times G(H_s)\sin\left(\pi\frac{\tau}{t_2^*}\right)dH_s(H_2>H_1;\ 0<t-t_1<t_2^*) \qquad (12)$$

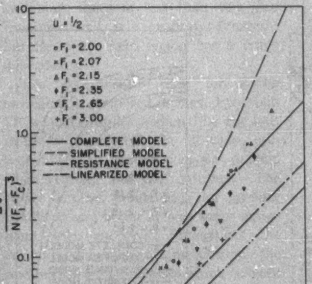


Fig. 4 (left). Discontinuity in voltage at t_1 versus normalized size of step, cubic dependence assumed, increased excitation

$$\frac{d\phi}{dt}=2B_sK_1K_2\int_{H_c}^{H_2}(H_2-H_s)G(H_s)\times \sin\left[\pi\left(\frac{t_1}{t_1^*}+\frac{t_1^*-t_1}{t_1^*\ t_2^*}\tau\right)\right]dH_s(H_2<H_1;\ 0<t-t_1<t_2^*) \qquad (13)$$

$K_1K_2G(H_s)$ is evaluated from constant-current switching data, and t_2^* is determined by requiring that the core saturate at $\tau=t_2^*$. Any other characteristics of this type of switching can then be calculated from these equations.

Linearized Form of Model

The complete model described by equation 6 is very difficult to handle mathematically in any case not involving constant-current excitation, and for this reason several simplifications have been suggested. One eliminates much of the complexity by considering all domains to be identical and thus characterized by the same starting field, the d-c coercive field.[7] For this representation the number of domains is unaffected by changes in excitation. To bring this into agreement with the actual dependence of the number of domains upon the excitation, the "constant" number of domains can be set to be consistent with the normal operating point, in the present case the initial level of excitation. With this compression of the starting field distribution, the equations describing the multidomain core are the same as those for any domain in it.

$$\frac{dr}{dt}=K(H-H_c) \qquad (14)$$

$$\frac{d\phi}{dt}=\pi B_sA_c(\sin\pi r)\frac{dr}{dt} \qquad (15)$$

which reduce to

$$i-i_c=\frac{el_c}{N^2B_sA_c\pi K\sin\pi r}=\frac{el_c}{N^2B_sA_c\pi K}\times \frac{1}{\sqrt{1-(1-2u)^2}}=\frac{e}{R(u)} \qquad (16)$$

where

l_c = mean length of core
N = number of turns in excitation winding
$u=\Delta\phi/\Delta\phi_{\max}=\Delta\phi/2B_sA_c$
K = constant chosen so that peak switching voltage is correct at the excitation level picked as a reference
$R(u)$ = apparent resistance of core
e = voltage across excitation winding

$$\frac{N}{l_c}(i-i_c)=H-H_c$$

This simplification of the model describes the unsaturated core as a current source shunted by a resistor dependent upon the instantaneous flux level, but not upon how that level was reached. The inherent disadvantage of this linearized system is that it only applies for limited deviations from the point where K was evaluated. The constant-current switching characteristics give K as a function of excitation, however, and the problem can be made piecewise linear if a very wide range of excitation is encountered.

Other Simplifications

The difference in waveshape between the induced voltage and the applied field makes analysis difficult for even slowly varying fields. It was suggested in reference 4 that the sinusoidal shape function might be replaced by its average value, and this procedure was shown to

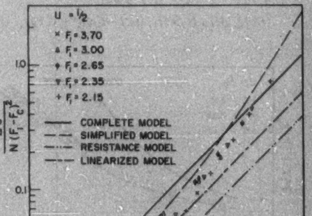


Fig. 5 (right). Discontinuity in voltage at t_1 versus normalized size of step, quadratic dependence assumed, increased excitation

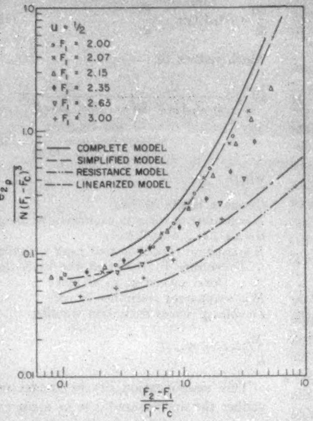
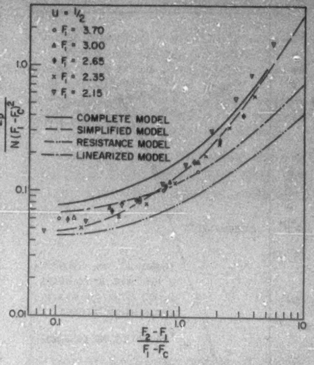


Fig. 6 (left). Peak switching voltage versus normalized size of step, cubic dependence assumed, increased excitation

Fig. 7 (right). Peak switching voltage versus normalized size of step, quadratic dependence assumed, increased excitation



give very good results when applied to a single-core magnetic-amplifier circuit. This representation still recognizes the distribution of starting fields, and the same type of correction as before must be applied to account for the reduction of potential nucleation sites by previous flux change. Thus, the describing equations become equation 17 for a single domain and equations 18, 19, and 20 for the entire reactor.

$$\frac{d\phi_1}{dt}=2B_sK_1(H-H_s)K_3\frac{2}{\pi} \qquad (17)$$

$$\frac{d\phi}{dt}=\frac{4}{\pi}B_sK_1K_3\int_{H_c}^{H_1}(H_1-H_s)G(H_s)dH_s \quad (0<t<t_1) \qquad (18)$$

$$\frac{d\phi}{dt}=\frac{4}{\pi}B_sK_1K_3\int_{H_c}^{H_1}(H_2-H_s)G(H_s)dH_s+\frac{4}{\pi}B_sK_1K_3\left(1-\frac{t_1}{t_1^*}\right)\int_{H_1}^{H_2}(H_2-H_s)\times G(H_s)dH_s(H_2>H_1;\ 0<t-t_1<t_2^*) \qquad (19)$$

$$\frac{d\phi}{dt}=\frac{4}{\pi}B_sK_1K_3\int_{H_c}^{H_2}(H_2-H_s)G(H_s)dH_s \quad (H_2<H_1;\ 0<t-t_1<t_2^*) \qquad (20)$$

This type of analysis should be expected to give good results only for predictions of total flux changes and similar noninstantaneous phenomena because the actual waveshape is ignored.

When the shape function in the linearized model is averaged, the result is identical with one of the most widely used models, the model of the core as a constant resistance shunted by a current source to provide the width of the direct-current hysteresis loop.[8-12]

$$i-i_c=\frac{el_c}{\frac{2}{\pi}N^2B_sA_c\pi K}=\frac{e}{\mathcal{R}} \qquad (21)$$

Once again this "constant" resistance must be determined for some nominal level of excitation; and as with the previous simplification, predictions of instantaneous quantities are expected to be in error.

Experimental Evaluation

The major point of interest in the work was the ability of the core models to correctly predict the operation of a reactor under double-step excitation. Accordingly, precautions were taken to eliminate extraneous effects even though they might normally be present in a magnetic-amplifier circuit. The inside-diameter to outside-diameter ratio of the core was chosen as 0.98 so that the excitation would be as uniform as possible. To reduce the capacitance and inductance of windings, very few widely spaced turns were used. When a core is brought

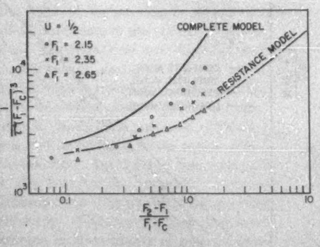


Fig. 8. Switching time after t_1 versus normalized size of step, cubic dependence assumed, increased excitation

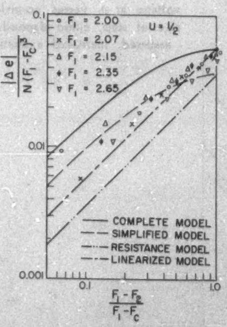


Fig. 9. Discontinuity in voltage at t_1 versus normalized size of step, cubic dependence assumed, decreased excitation

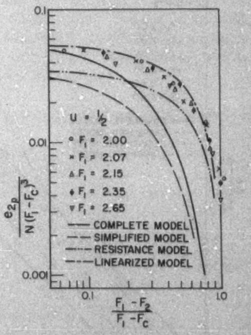


Fig. 10. Peak switching voltage versus normalized size of step, cubic dependence assumed decreased excitation

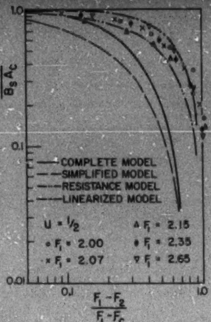


Fig. 11. Flux change between t_1 and t^* versus normalized size of step, cubic dependence assumed, decreased excitation

rapidly out of saturation, a short-duration voltage spike is invariably produced, usually attributed to reversible domain rotation. This spike is not accounted for by any model. While it may not represent useful flux change in the core producing it, it is effective in the control of any other resetting core. For this reason no attempt was made to perform tests in a multicore circuit or to use cores for excitation purposes. The circuit used is shown in Fig. 2. As illustrated, it will produce increases in excitation. The 8-turn winding is reversed for decreases. The constant-current switching characteristic, required to evaluate certain constants of the model, is shown in Fig. 3. This is not the quadratic characteristic reported in reference 4, but rather one with two distinct portions. The higher field portion, applying for switching times of less than 900 μsec (microseconds), is a quadratic function; but the lower field region has a cubic dependence.

The experimental data for 2-step switching is compared in Figs. 4 through 11, with the results predicted by the four methods considered. In all cases presented here the second step was applied when the core was half switched. The choice of other points does not produce results in any better or any worse agreement with theory.[13] Because of the extreme difficulty of performing the mathematics for processes which covered both regions of the constant-current switching characteristic, experimental results are compared with the predictions of the mode for processes restricted to one region. If this limitation is not observed, rather great deviations can be produced, as is shown in several of the illustrations. The features of the switching behavior which are shown are the discontinuity in induced voltage at the time of application of the second step and the peak switching voltage reached. In addition, the time required to reach saturation after the second step is shown for cases of increased excitation; and the change in flux between the beginning of the second step and the originally expected time of saturation is shown for decreases in excitation. For completeness, Figs. 12 and 13 show the type of deviation produced by using the complete experimental constant-current switching characteristic in the complete model. These predicted results cannot be normalized, and therefore, only one magnitude of initial excitation is considered.

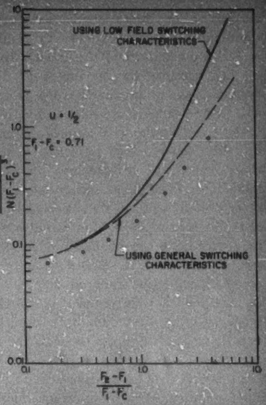


Fig. 12. Peak switching voltage versus normalized size of step, general and low field distribution, increased excitation

Conclusions

As is evident from the illustrations, neither the complex complete model nor the simplified versions give accurate enough predictions of the behavior of a reactor with discontinuously increasing excitation to make their use advisable for quantitative purposes. The errors are large in themselves and the positive feedback inherent in many multicore circuits would be expected to produce still worse agreement.

Certain qualitative features do appear that are worth mention as a guide to future work. The induced-voltage waveshapes, predicted for excitation increases by the complete model and by one not considering a modified shape function, are shown in Fig. 14. The observed waveshape was of the first type, the late-occurring peak produced by the modified shape function being apparent even when the core was three-quarters switched at the time of the second step. Models attempting to relate flux, rate of change of flux, and magnetomotive force without a knowledge of the past excitation history of the core cannot take this into

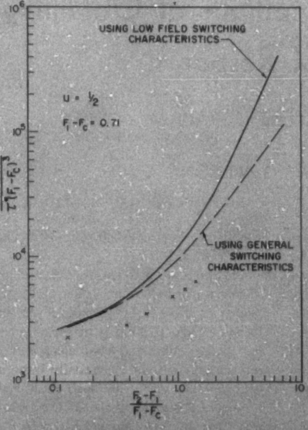


Fig. 13. Switching time after t_1 versus normalized size of step, general and low field distribution, increased excitation

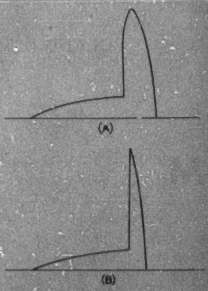


Fig. 14. Predicted switching voltage waveforms

A—Shape function and number of domains dependent on past
B—Only number of domains dependent on past

account. For excitation increases the discrepancy between experimental results and the predictions of the models that do not consider the starting field distribution gets progressively worse as the size of the step is increased. Further, if the complete model is applied neglecting the domain nuclei reduction factor, the disagreement is increased. Thus, a model to handle excitation increases must probably utilize these two features of the present model. The discontinuity in induced voltage at the time of the second step should, according to the complete model, be produced only by the domains which were growing under the influence of the first step, and should be linearly dependent upon the increase in excitation. While this is the case for large initial excitations, it is not so for smaller initial excitations, which would indicate that the effective starting field is a function of domain size or excitation. The other alternative, that the discontinuity is dependent to a considerable extent upon domains which have just started to grow, would require a similar explanation.

For cases of discontinuously decreased excitation, it is possible to predict the core behavior fairly accurately, but not by using the complete model. That is very inaccurate. The best results are obtained with the linearized model, and the next best with the resistive model. These models do not consider the starting field distribution and therefore do not alter the number of growing domains when excitation is decreased. These two models also have the property that flux, rate of change of flux, and magnetomotive force are uniquely related, or at least do not depend upon past-excitation, so that extensions of experimental data using only these three quantities is justified for decreases in excitation.[5] This behavior again supports the idea that the effective domain starting field must vary with domain size or with excitation. A similar behavior has been reported for *65* permalloy but the explanation presented in that case involved surface phenomena.[14] These could not be incorporated in the present model without extensive revision. Here domain nuclei were assumed distributed throughout the volume of the material, and no special significance was attached to those at the surface.

References

1. Magnetic Materials for Digital Computer Components: I, A Theory of Flux Reversal in Polycrystalline Ferromagnetics, N. Menyuk, J. Goodenough. *The Physical Review*, New York, N. Y., vol. 97, 1955, pp. 8–18.

2. Terminal Properties of Magnetic Cores, T. C. Chen, A. Papoulis. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 46, 1958, pp. 839–49.

3. Some Switching Properties of Square Loop Ferrites, E. M. Gyorgy, J. L. Rogers. "Conference on Magnetism and Magnetic Materials," *AIEE Special Publication T-91*, Feb. 1957, pp. 687–40.

4. A Mathematical Model for a Magnetic-Amplifier Reactor Core, H. H. Woodson. *AIEE Transactions*, see pp. 872–78 of this issue.

5. Dynamic Core Behavior and Magnetic Amplifier Performance, L. A. Finzi, D. L. Critchlow. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, May 1957, pp. 229–40.

6. Core-Reset Functions in Magnetic-Amplifier Analysis, G. C. Feth. *Ibid.*, vol. 77, Sept. 1958, pp. 503–19.

7. Analysis of a Three-Phase Magnetic Amplifier with Resistive Load, J. Borrego-Larralde. *S.M. Thesis*, Massachusetts Institute of Technology, Cambridge, Mass., 1956.

8. The Influence of Eddy Currents Upon the Performance of A-C and D-C Magnetic Amplifiers, R. Zarouni. *Report PIB-227*, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., 1952.

9. A Note on the Behavior of an Idealized Half-Wave Magnetic Amplifier in the Presence of Eddy Currents, E. Smith. *Report PIB-267*, Polytechnic Institute of Brooklyn, 1953.

10. On the Mechanics of Self-Saturating Magnetic Amplifier Operation, T. Kusada. *Journal of the Institute of Polytechnics*, Osaka City University, Osaka, Japan, series E, vol. 3, 1956, pp. 103–18.

11. Analysis of the Full-Wave Magnetic Amplifier Circuits Considering the Change of the Width of the Dynamic Hysteresis Loop, Tadashi Kikuchi. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, July 1957, pp. 241–8.

12. Analysis of Magnetic Amplifiers Using a Resistive Reactor Model, Shimshon Frankenthal. *Ibid.*, see pp. 850–57 of this issue.

13. The Validity of Present Core Models for the Analysis of Multi-core Magnetic Amplifiers, B. Lovell. *S.M. Thesis*, Massachusetts Institute of Technology, 1958.

14. Some Properties of the Coercive Force in Soft Magnetic Materials, D. Rodbell, C. Bean. *The Physical Review*, vol. 103, 1956, pp. 886–95.

Some Practical Methods of Obtaining Time of Response

ROLAND L. VAN ALLEN
ASSOCIATE MEMBER AIEE

PAUL W. COVERT
ASSOCIATE MEMBER AIEE

THOMAS G. WILSON
ASSOCIATE MEMBER AIEE

MEASUREMENTS of response time of magnetic amplifiers and similar carrier-modulated control devices have always presented problems to the design engineer. A number of methods have been described in the literature. Some of these have been proved to be reliable and accurate. Nevertheless, practical considerations often have limited their widespread usage.

This paper presents a brief description of three methods of measuring time of response to a step function. They have been found to be particularly useful because of their simplicity and convenience.

The first method entails the use of a "firing angle indicator." The indicating device itself is simply a small pulse transformer. The technique involved is similar to that employed when using a cyclic integrator. Instead of actually integrating the output over each half-cycle, however, a simple phase-shift circuit is used to produce an output which is equivalent to that which would be obtained from a magnetic amplifier having certain ideal characteristics. The time from initiation of the step-function transient until the output reaches a fixed percentage of its excursion is recorded and is the response time of the amplifier.

The second method, referred to as the "core integration" method, utilizes a toroidal core with a square-hysteresis-loop characteristic. The number of turns on this core establishes a fixed volt-second reference. The core is excited by the output voltage from the magnetic amplifier. Saturation of the core indicates when the volt-second integral of the output exceeds the reference level established by the wound core. A record is made of the time from the initiation of a transient

---

Paper 59-1061, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-IRE Nonlinear Magnetics and Magnetic Amplifiers Conference, Washington, D. C., September 23–25, 1959. Manuscript submitted June 24, 1959; made available for printing July 16, 1959.

Paul W. Covert is with Magnetics, Inc., Butler, Pa. Roland L. Van Allen, National Aeronautics and Space Administration, Washington, D. C., and Thomas G. Wilson, Duke University, Durham, N. C., are both formerly of Magnetics, Inc., Butler, Pa.

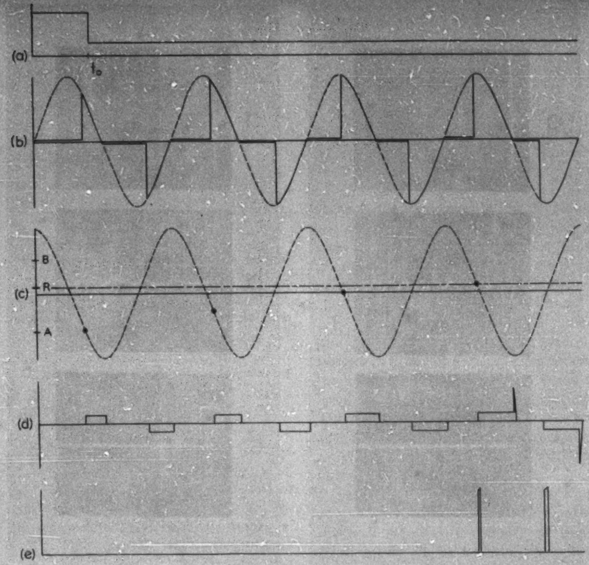


Fig. 1. Instantaneous waveshapes during a transient

a—control-circuit voltage
b—amplifier load voltage
c—output of firing angle indicator circuit with the reference cosine wave. A, minimum test output level; B, maximum test output level; R, 63%-reference level
d—output of core integration circuit
e—output of comparison circuit

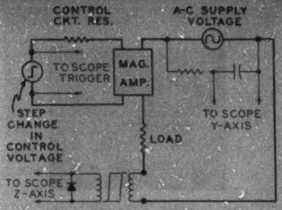


Fig. 2. Circuit for firing angle indicator method

until the average output of the amplifier reaches the preset value of the reference.

The third method describes a "comparison" in which an auxiliary magnetic amplifier is used to establish a reference at some known output level. An indication is made during a transient when the magnetic amplifier under test delivers an output equal to that established by the reference amplifier. Measurements of time of response are made by recording the time from initiation of a step function until the output of the amplifier under test equals that of the reference.

Considering just the variety and number of different approaches that have been made in solving this problem in measurements, there is evidence that the single ideal solution applicable to all situations has not been fully resolved. The particular areas of usefulness for each of the three methods along with their advantages and disadvantages are discussed.

Background

The AIEE Magnetic Amplifier Committee has given consideration to the problem of measuring the response time of magnetic amplifiers and has proposed test codes for the express purpose. Section V.B.1 of the proposed Standard Test Codes[1] states the following concerning the measurement of response time. "Determine the time (preferably in seconds; may also be measured in cycles of the power frequency) required after application of the step-function signal for the output quantity to change by an amount equal to 63% (unless otherwise specified) of the final change in output." For nonreversible magnetic amplifiers, the final change in output has been proposed as being from 20% to 80% of the value between minimum and rated output.

With regard to the nature of the output quantity itself, the proposed standard recognizes that the intended use of the amplifier often dictates that characteristic which will be of greatest interest. This may be either the half-cycle average output voltage,[2] the rms output, or the fundamental component of the output. Whatever it is, a conversion chart may be made by steady-state measurements to permit translation to and from the parameter actually sensed or detected by the measuring device.

It should be noted that the AIEE Magnetic Amplifier Committee has only indicated what is to be measured and has not specified how the measurement is to be made. Most of the techniques or methods indicating how to measure response time depend upon either performing the half-cycle by half-cycle integration of the output voltage or else singling out and identifying the particular half-cycle when the output passes through the reference 63% point.

The use of a cyclic integrator[3,4] is no doubt one of the best methods to display the dynamic behavior of a magnetic amplifier since it can measure the average output each half-cycle. From a practical viewpoint, however, it requires special electronic circuits. This fact probably has retarded its general acceptance by industry. A different method which has found wide usage measures the variation in firing angle using an oscilloscope.[5] This latter method, while convenient and simple for relatively high-speed magnetic amplifiers, is not suitable for measuring long response times. For these and other reasons, need has arisen for methods which, from a purely practical standpoint, could be used without requiring elaborate preparation.

Time of Response With Firing Angle Indicator

THEORY

With reference to Fig. 1, (b) is a sketch of the output load voltage $e_L(wt)$ of a magnetic amplifier which might occur

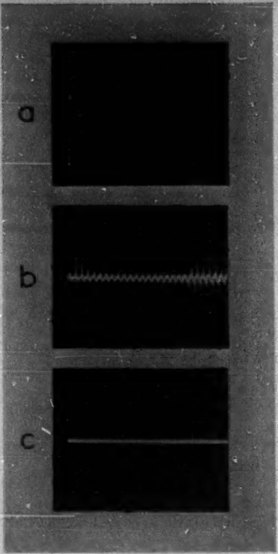


Fig. 3. Oscillograms of output with N_c^2/R_c = 1,620 using sweep of 50 milliseconds division

a—firing angle indicator circuit
b—core integration circuit
c—comparison circuit

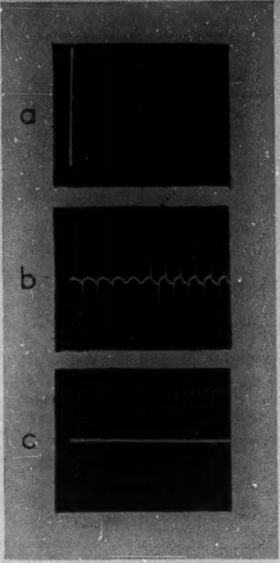


Fig. 4. Oscillograms of output with N_c^2/R_c = 270 using sweep of 20 milliseconds division

a—firing angle indicator circuit
b—core integration circuit
c—comparison circuit

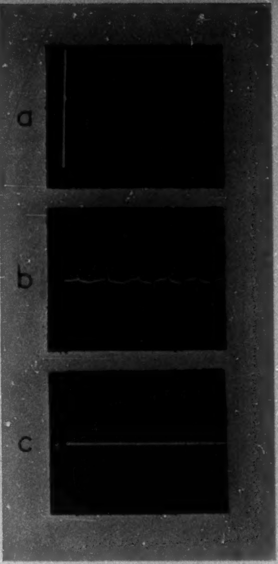


Fig. 5. Oscillograms of output with N_c^2/R_c = 81 using sweep of 10 milliseconds division

a—firing angle indicator circuit
b—core integration circuit
c—comparison circuit

in response to a transient initiated by a step change in control-circuit voltage. The latter is shown by the sketch (a). For the present, focusing attention only on (a) and (b), the average output during each half-cycle of the transient is obtained by integration. Calling E_{a+} the output of the plus half-cycles and E_{a-} that of the negative half-cycles, the expressions are

$$E_{a+} = \frac{1}{\pi}\int_0^{\pi} e_L(wt)d(wt) \qquad (1)$$

$$E_{a-} = \frac{1}{\pi}\int_{\pi}^{2\pi} e_L(wt)d(wt) \qquad (2)$$

The true cyclic integrator performs these computations each half-cycle and usually produces an output pulse at the end of each half-cycle which has an amplitude proportional to E_{a+} and E_{a-}.

If the load on the amplifier is resistive and if magnetizing current can be neglected, then the assumption that the load voltage approximates a portion of a sinusoid permits some simplification. Considering just the positive half-cycle and calling the firing angle during this period α_+, equation 1 reduces as follows:

$$E_{a+} = \frac{1}{\pi}\int_{\alpha_+}^{\pi} \hat{E}_L \sin(wt)d(wt) = \frac{\hat{E}_L}{\pi}(1+\cos\alpha_+) \qquad (3)$$

This expression indicates that even during a transient when the angle α_+ is changing from cycle to cycle, the half-cycle average of the output can be expressed as a function of the cosine of the angle at which firing took place. This fact is central to the first method of measurement.

METHOD OF MEASUREMENT

Fig. 2 shows symbolically a magnetic amplifier with an a-c load. An R-C (resistance-capacitance) phase shifting network to generate a cosine wave is located across the a-c voltage supplying the amplifier. The primary of a small pulse transformer is in series with the load. This transformer employs a square-loop core material and is designed to support a voltage which is small with respect to the gate circuit supply. It saturates in a few millivolt-seconds and its output is used to modulate the intensity of the beam of the oscilloscope. The design is such that it will intensify the beam for only a few microseconds at the instant the magnetic amplifier fires. The silicon diode placed across the secondary of the transformer permits only a unidirectional pulse to the Z-axis of the oscilloscope and eliminates ringing in the circuit.

With this configuration, there will be a dot appearing on the face of the oscilloscope once each positive half-cycle. Its vertical position will be equal to the instantaneous value of the cosine wave at the moment the magnetic amplifier fires, or proportional to cos α_+.

To obtain transient response, a step-function change in the control-circuit voltage supplying the amplifier is made. This signal is used to trigger the sweep circuit of the oscilloscope. The sweep of the oscilloscope serves as the time base for measuring response time.

Fig. 1(c) shows, as a dashed curve the cosine wave impressed on the Y-axis. The intensity of the beam, however, is adjusted so that nothing is visible unless a d-c potential exists at the Z-axis terminals. This occurs only at the instant of firing and is indicated by the heavy

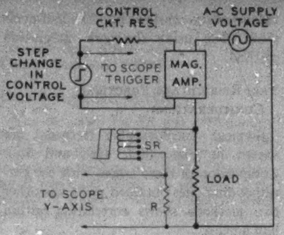


Fig. 6. Circuit for core integration method

dots. A dot at the negative peak of the cosine wave ideally would indicate zero output and one at the positive peak would indicate the theoretical maximum.

Results of measurements of time of response using the firing angle indicator are shown in Figs. 3(a), 4(a), and 5(a) along with comparative results obtained using the other two methods. These figures show the effect of a decreasing N_c^2/R_c on a test amplifier. Details of the experimental circuits used in obtaining these data are given in the Appendix. All oscillograms are the result of "rise-time" measurements. The traces labeled (a) in these figures show, on the left, a reference dot at a level corresponding to a 63% change in output. The time of response in seconds obtained by this method is measured from the calibrated sweep of the oscilloscope from time t_0 until the output reaches the 63%-reference level. Time of response in cycles may be obtained by simply counting the dots, since one dot occurs each cycle.

When fall-times are measured, the appropriate 63% change in output is recorded for reference. It should be recognized that the rise-time indicated using this method will always be slightly shorter than the fall-time. This is caused by the propagation of the firing angle which advances toward the beginning of the half-cycle when rise-time measurements are made and toward the end of the half-cycle when fall-time measurements are made. For most practical measurements, this small error may be neglected as accuracy of ±1/2 cycle is satisfactory.

This characteristic inherent in this method of measurement brings out a basic difference between the "firing angle indicator" and the cyclic integrator. The cyclic integrator indicates the output at the end of each half-cycle, whereas the firing angle indicator indicates the level of output at the instant the amplifier fires. Further, with the firing angle indicator, measurements are made only once each cycle which is a disadvantage with 2 cycles or less response time.

Time of Response by Core Integration

CIRCUIT DESCRIPTION

The same sudden saturation characteristic utilized in the design of the saturable reactors for high-performance magnetic amplifiers also can be employed to measure their response time. This is illustrated by the circuit of Fig. 6. Here, an auxiliary circuit, consisting of a saturable reactor *SR* and a resistor *R*, is connected in parallel with the a-c load on the amplifier. The voltage across *R* is fed to the *Y*-axis of an oscilloscope.

In the design of this auxiliary circuit, *R* is chosen to be large enough so that even if *SR* were short-circuited, the current drain would not exceed a few per cent of the amplifier load current. The other component, saturable reactor *SR*, is constructed from a highly rectangular-loop core material, such as Orthonol, and has numerous taps.

The circuit functions in the following way. Immediately after the firing of the magnetic amplifier, a voltage essentially equal to the entire load voltage $e_L(wt)$ appears across *SR* and causes its flux to change. The only voltage appearing at the *Y*-axis terminals during this period is the *IR* drop due to the magnetizing current for *SR*. If the load voltage becomes high enough, however, *SR* can no longer absorb all of the voltage, and it saturates. Coincident with saturation, the voltage across *R* suddenly increases becoming nearly equal to the load voltage itself.

To measure time of response using the circuit of Fig. 6, the number of turns *N* on *SR* are selected so that the reactor is just on the verge of saturating when the output load voltage of the amplifier is at the 63%- point described earlier. With *N* so selected, the load voltage is set to the "minimum test output voltage"[1] and a step-function of control is applied to take the output to the "maximum test output voltage." A sketch of the wave-shape occurring across *R* is shown in Fig. 2(d). The response time is that period of time from t_0 until the first spike of voltage occurs across *R* indicating saturation of *SR* and showing that the amplifier has completed 63% of its excursion.

In comparing Fig. 1(c) and 1(d), it should be noticed that in the core integration method, indication that the amplifier has reached the prescribed reference value occurs near the end of the half-cycle. On the other hand, in the firing angle indicator method, indication is given near the beginning of the half-cycle in which the output will first exceed the reference value. This difference should be borne in mind when comparing results from the two methods.

MEASUREMENT RESULTS

Figs. 3(b), 4(b), and 5(b) show oscillograms of the voltage appearing across the resistor *R* of Fig. 6 during increasing transients of load voltage. The pictures show unidirectional pulses of output occurring during the first few cycles of the transient. They are produced because the flux level in *SR* at the instant the transient is initiated is arbitrary and may

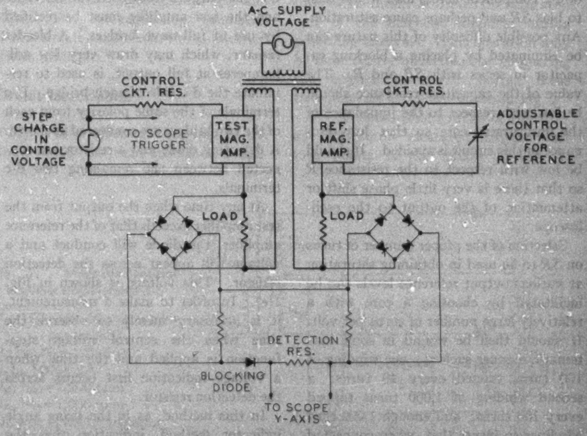


Fig. 7. Circuit for comparison method

Table I. Recommended Use of the Three Methods for Measuring Time of Response

	Firing Angle Indicator Method	Core Integration Method	Comparison Method
Use on slow transients	Good	Good	Good
Use on fast transients	Over 2 cycles	Over 2 cycles	Good
Usable with high-frequency a-c supplies	Requires core with fast switching time	Yes	Yes
Usable for a-c resistive loads	All except high *R* values	All values	All values
Usable for d-c inductive loads	No	Yes	Yes
Usable for d-c loads	Only with modifications	Series capacitor required	Yes
Usable for reversible output	Not practical	Series capacitor required	Yes

be at any point between $+B_r$ and $-B_r$. If the flux should have been left near one of the residual points, then the core probably would saturate in this direction if the load voltage during the first half-cycle of the transient should be of such a polarity to drive it toward this residual point. These pulses disappear within a few cycles, but care should be used in interpreting oscillograms showing fast transients so as not to confuse these pulses with the bi-directional pulse which occurs when the 63%- point is passed. Extrapolation of the envelope of the bi-directional pulses back to zero often helps in locating this point.

PRACTICAL CONSIDERATIONS

This measurement approach is based on the integrating property of square-loop core materials and requires that the voltage impressed upon the *R-SR* circuit be a-c. Sometimes it happens that the amplifier load voltage contains a small d-c component. This might be due to a slight unbalance between the two gate windings or the two rectifier characteristics. Whatever the origin, the existence of d-c component across load would tend to bias *SR* and perhaps cause saturation. Any possible difficulty of this nature can be eliminated by placing a blocking capacitor in series with *SR* and *R*. The value of the capacitive reactance should be low with respect to the impedance of the unsaturated core so that ferroresonance in this circuit is avoided. It should be low with respect to the resistance *R* so that there is very little phase shift or attenuation of the output to the oscilloscope.

Selection of the proper number of turns on *SR* to be used in obtaining saturation at various output reference levels can be facilitated by choosing a core with a relatively large number of turns per volt. It should then be wound in some systematic manner such as: one winding of 100 turns tapped every 10 turns; a second winding of 1,000 turns tapped every 100 turns; and enough 1,000-turn windings to insure that, when connected in series at the highest reference voltage anticipated, the core is saturated. Selection of the proper number of turns to be connected in series then becomes an accurate and rather rapid process.

Time of Response by Comparison Method

MECHANICS OF OPERATION

This technique of measurement is referred to as the comparative method since a reference amplifier, having similar characteristics to the amplifier being measured, is needed to pre-establish an output level against which the output from the test amplifier may be compared.

The mechanics of operation can be understood by considering the circuit of Fig. 7. The reference amplifier is adjusted to some arbitrary output level which normally blocks a simple diode detector. When the output from the test amplifier just exceeds the output from the reference, current flows through the detection branch.

Both amplifiers should have their rated loads connected as in normal operation Unless the outputs are already d-c, the outputs from both the reference and the test amplifier must be rectified by use of full-wave bridges. A bleeder resistor, which may draw very few milliamperes at full output, is used to terminate the d-c side of each bridge. D-c terminals of the same polarity from each of the two bridges are connected as shown. A detecting diode and a resistor are connected between the remaining two d-c terminals.

At any time when the output from the test amplifier exceeds that of the reference amplifier, the diode will conduct and a voltage will appear across the detection resistor. This voltage is shown in Fig. 1(e). In order to make a measurement, it is necessary merely to observe the time when the control voltage step-function is applied and the time when a voltage indication first occurs across the detection resistor.

In this method, as in the firing angle indicator method, indication that the output has arrived at the preset reference value occurs earlier in the half-cycle than when using the core integration method.

TEST RESULTS AND PRACTICAL CONSIDERATIONS

Typical oscillographic records are shown in Figs. 3(c), 4(c), and 5(c). Comparison of these traces with the corresponding ones obtained using the other two methods shows correlation within half a cycle.

A simple modification to the detection branch makes it quite easy to use this scheme to measure response time of reversible-output magnetic amplifiers. This has been employed successfully on a production line where time of response of such amplifiers had to be continually checked and recorded.

The comparison approach is particularly adaptable when measurements are required on amplifiers whose load waveshapes are quite distorted. The only requirement is that the reference amplifier be similarly terminated.

Conclusions

The three methods presented for measuring time of response have certain features in common. None requires special equipment, elaborate preparation time, or a skilled operator. All are accurately reproducible and widely applicable.

Table I shows a comparison of the three methods. It can be used as a guide in selecting the most suitable method to be applied when special measurement problems are encountered.

Appendix

Magnetic Amplifier Circuit

All oscillograms were taken using a doubler-connected magnetic amplifier with a-c output. The amplifier cores were Magnetics, Inc. *50,103-4A* (2 × 3 × 1/2 inch) wound with gate windings of 2,350 turns and a control winding of 900 turns. The self-saturating diodes were silicon, Westinghouse type *3080*. The load resistance was 100 ohms.

All tests were conducted at 120 volts, 60 cps (cycles per second). Under this condition, the "rated output voltage" was taken as 85 volts for the half-cycle average output. The "minimum output voltage" was measured as 0.1 volt. This yields 17 volts for the "minimum test output voltage" and 68 volts for the maximum test output voltage." The 63%- point for a rising transient is thus 49.1 volts on a half-cycle average bias.

Firing Angle Indicator Circuit

The pulse transformer used to record the data presented was made with a Magnetics, Inc., core *50,000-4A* (1/2 × 3/4 × 1/4 inch)

Vol.
Pt. 1
1959

wound with 300 turns on the primary and 100 turns on the secondary. The diode in the secondary circuit was a GE (General Electric Company) *1N538*.

Core Integration Circuit

The core used for *SR* was a Magnetics, Inc., *50,026-2A* (1 × 1½ × 3/8 inch) with various windings. The turns actually used in recording the data were 2,620. The resistor *R* was 5,100 ohms. In all tests, a capacitor of 4 microfarads was used in series with *SR* and *R*.

Comparison Circuit

The reference amplifier was identical to the test amplifier. The auxiliary bridge rectifiers paralleling the two 100-ohm loads were made using *GE 1N538* silicon diodes. The bleeder resistors were 10,000 ohms, and the detection resistor 10,000 ohms. The blocking diode was a *GE 1N538*.

References

1. Proposed Standard Test Codes for Magnetic Amplifiers, AIEE Committee Report. *AIEE Transactions*, pt, I. (*Communication and Electronics*), vol. 78, Sept. 1959, pp. 453–56.

2. Recommended Symbols for Magnetic Amplifier Papers—A Report, AIEE Committee Report. *Ibid.*, Nov. 1959, p. 520, Note C1.

3. A Transient Analyser for Magnetic Amplifiers, E. J. Smith. *Ibid.* vol. 72, Sept. 1953, pp. 461–65.

4. The Cyclic Integrator—A Device for Measuring the Frequency Response of Magnetic Amplifiers, T. Dunnegan, Jr., J. D. Harnden, Jr. *Ibid.*, vol. 73, Sept. 1954, pp. 358–65.

5. Oscillographic Techniques for the Evaluation of Magnetic Amplifier Response, D. L. Critchlow. *Ibid.*, vol. 74, Nov. 1955, pp. 607–10.

Self-Regulation in Magnetic-Transistor Amplifiers

C. E. HARDIES
ASSOCIATE MEMBER AIEE

R. L. VAN ALLEN
ASSOCIATE MEMBER AIEE

MAGNETIC amplifiers incorporating transistors have received considerable emphasis in the literature. Many circuits have been devised using magnetic amplifiers and transistors in combination to achieve many different characteristics. This paper describes a characteristic available with this combination of devices that has not been heretofore described. By utilization of the circuitry to be described, an amplifier can be designed that has a self-regulated output for variations in supply excitation. This characteristic is inherent to the circuitry; not accomplished by the addition of compensating networks. While self-regulation is the primary advantage to be gained by use of the circuitry, no sacrifice is made in gain, temperature stability, or frequency stability from that of a conventional magnetic amplifier. In addition, since the transistor is utilized only as a switch (transistors have very low saturated resistances as compared with magnetic amplifiers) the regulation of the device with load is very good.

The basic method to be discussed for achieving self-regulation for variations in supply excitation used in the present device has been treated in all-magnetic devices called the "transactor"[1] and the "shunt-coupled magnetic amplifier."[2]

In these devices, the output is taken essentially in parallel with the magnetic cores so that output exists during the time interval while the cores are changing flux. This parallel operation requires that current limiting impedances be placed in series with the gate windings to limit the gate current after firing. Actually, the principle of having output occur during the prefiring interval rather than in postfiring intervals, as with conventional magnetic amplifiers, is the only item common to these all-magnetic devices and the magnetic-transistor amplifier. The self-regulating properties of this circuit result from the application of this principle.

Circuit Description

Fig. 1 is a schematic diagram of the single-ended magnetic-transistor amplifier. It contains a basic center-tapped magnetic amplifier and a transistor, together with their associated power and bias supplies.

The polarity of the transistor bias, acting through R_b, is negative to the base of the transistor and sufficient in magnitude to hold the transistor in its "on" or short circuit condition under the most adverse conditions of load, excitation, and temperature. The polarity of the magnetic amplifier is such that its output is positive with respect to the base of the transistor. Thus, output is delivered via the switched transistor during the time intervals before the magnetic amplifier fires (prefiring), and and no output is delivered during the time intervals either core is saturated (postfiring). The transistor is held "off" or open circuited during postfiring intervals because of the positive voltage established at the base by the magnetic amplifier. The bias current which passed through the base during the prefiring intervals is shunted through the saturated core of the magnetic amplifier during postfiring intervals. This operation may be more clearly understood by comparing the voltage waveforms during these two intervals which are shown in Fig. 2.

Fig. 2 shows photographs of the voltage waveforms measured between various points in the circuit of Fig. 1.

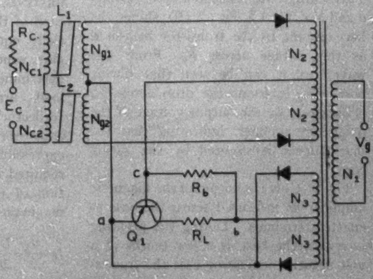


Fig. 1. Magnetic-transistor amplifier single-ended d-c

V_g, 120 volts 60 cps (cycles per second)
N_1: N_2: N_3 = 120: 6: 30
L_1, L_2, 50004-2A
N_{g1}, N_{g2} = 1,200 turns
N_{c1}, N_{c2} = 1,000 turns
Q_1, type 2N241A transistor
R_b = 2,000 ohms
R_L = 150 ohms

Paper 59-1059, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-IRE Nonlinear Magnetics and Magnetic Amplifiers Conference, Washington, D. C., September 23–25, 1959. Manuscript submitted June 24, 1959; made available for printing July 16, 1959.

C. E. Hardies is with Magnetics, Inc., Butler, Pa. R. L. Van Allen, formerly with Magnetics, Inc., is now with National Aeronautics Space Administration, Washington, D. C.

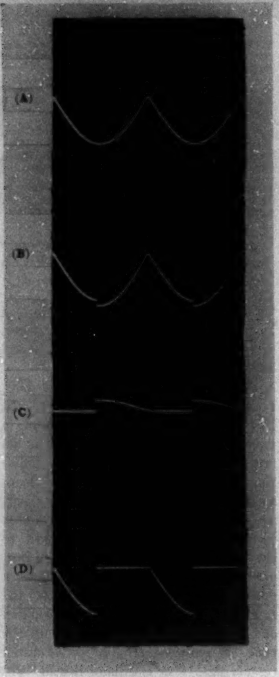


Fig. 2. **Oscillograms of voltages from circuit of Fig. 1**

A—Voltage a to b
B—Voltage across R_b
C—Voltage a to c
D—Voltage across R_1

The signal level is such that the magnetic amplifier fires at approximately 90 degrees. Each of the photographs show one full cycle of the circuit operation. Fig. 2 (A) shows the power supply voltage to the transistor and load. This voltage was measured across points *a* and *b* of Fig 1. Fig. 2 (B) shows the bias current to the transistor measured as the voltage across R_b. From this waveform, it can be seen that during post-firing intervals the drop across R_b is increased by the output voltage of the magnetic amplifier, indicating that the bias current is shunted by the gating core.

Fig. 2 (C) is the output of the magnetic amplifier as measured across the base to emitter terminals. During the prefiring intervals, the base is a few tenths of a volt negative indicating that the transistor is saturated or short circuited. During the postfiring intervals this voltage is positive, effectively biasing the transistor off.

Fig. 2 (D) is the waveform of the voltage measured across load resistor R_L. As has been described, the output occurs during the prefiring intervals. This output has a constant volt-second value, determined by the reset of the magnetic cores. The output of the transistor occurs in the prefiring interval determined by the firing angle of the magnetic amplifier. Then, like the transactor and shunt-coupled amplifier, the output is self-regulating with variations in supply voltage. When the supply voltage is increased the instantaneous value of the output will increase, but the firing angle will occur earlier in time maintaining constant volt-seconds and thus regulating the output. Likewise, when the supply voltage is decreased, the firing angle will occur later in time with lower instantaneous values also regulating the output.

Fig. 3, curve A shows a typical transfer characteristic for the amplifier shown schematically in Fig. 1. The magnetic-transistor device has zero output at zero control signal, because with the self-saturating magnetic amplifier full on, the transistor is biased off. This transfer characteristic was taken from a unit containing 50-50 nickel-iron magnetic cores having a high squareness ratio.

Design Considerations

The magnetic amplifier in the magnetic-transistor amplifier shown in Fig. 1 operates in the conventional manner with the base circuit of the transistor as its load. The base diode conducts bias current in its forward direction during prefiring intervals. During this time, the magnetic amplifier's load can can be considered as being made up of a resistor and battery in series. The battery voltage being the forward drop across the base diode. In post-firing intervals, the current which will flow in each of the gate windings will be essentially equal to the bias current. The gate windings current capacity can be determined directly from the required load current since the bias and load (collector) currents are related in the transistor by its gain. The following expression yields the current rating required in the gate windings as a function of the load current of the entire magnetic-transistor device.

$$I_g = \frac{I_b}{2} = \frac{I_c}{2\beta}$$

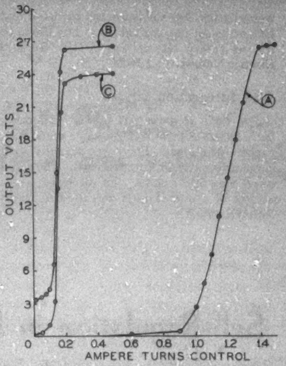


Fig. 3. **Transfer characteristic for single-ended d-c magnetic transistor amplifier**

A—With L_1, L_2 made from Orthonol
B—With L_1, L_2 made from HyMu80
C—With L_1, L_2 made from HyMu80 and with circuit off

where

I_g = average current in the gate windings of the magnetic amplifier
I_b = average bias current
I_c = collector and average current through the load R_L
β = current gain of the transistor, grounded emitter

The voltage output of the magnetic amplifier is dictated by the transistor base diode. The output voltage from the magnetic amplifier must be kept below the ratings for maximum reverse voltage on the base to emitter junction.

With these current and voltage requirements for the magnetic amplifier's load circuit, one can design a magnetic amplifier to suit the other requirements of the particular application such as gain, sensitivity, and response time according to conventional magnetic amplifier techniques.

The transistor used in the magnetic-transistor amplifier is selected so as to be able to withstand the required supply voltage when it is open-circuited and pass the load current when fully turned on. In cases where the maximum collector current is not specified by the manufacturer, this can be determined from the saturated resistance of the device and the allowed junction dissipation. In determining collector current, the transistor must be derated for the highest ambient temperature in which the device is to be used.

The value of the bias resistor is selected

Vol.
Pt. 1
1959

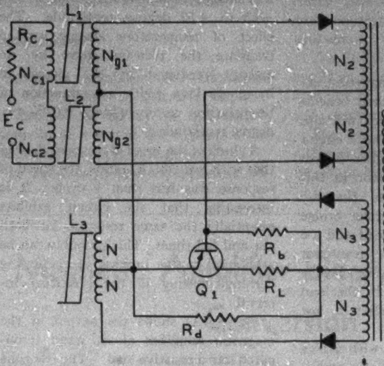

L1
Rc
Nc1
Ng1
L2
Ec
Nc2
Ng2
L3
N
N
Q1
Rb
RL
Rd
N2
N2
N3
N3
N1
Vg

Fig. 4. D-C, magnetic-transistor amplifier using sensitive core material

V_g, 120 volts 60 cps
N_1: N_2: N_3 = 120: 3: 30
L_1, L_2 50004-2D
N_{g1}, N_{g2} = 1,200 turns
L_3, 50007-2A
N = 600 turns
N_{c1}, N_{c2} = 1,000 turns
Q_1, type 2N241A transistor
R_b = 2,000 ohms
R_L = 150 ohms
R_d = 10,000 ohms

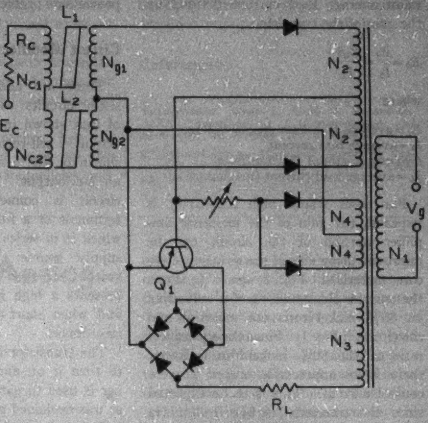

L1
Rc
Nc1
Ng1
L2
Ec
Nc2
Ng2
Q1
N2
N2
N4
N4
N3
N1
Vg
RL

Fig. 5. A-C, magnetic-transistor amplifier

V_g, 120 volts 60 cps
N_1: N_2: N_3: N_4 = 120: 6: 30: 30
L_1, L_2, 50004-2A
N_{g1}, N_{g2} = 1,200 turns
N_{c1}, N_{c2} = 1,000 turns
Q_1, type 2N241A transistor
R_b = 2,000 ohms
R_1 = 150 ohms

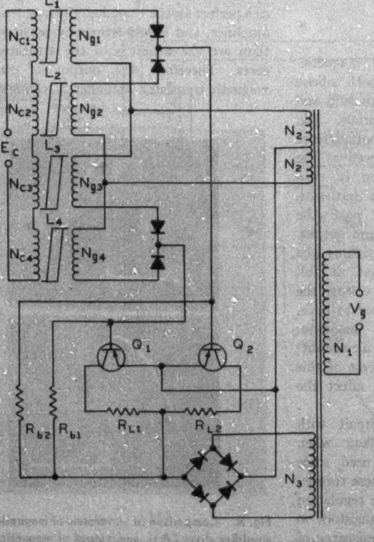

L1
Nc1
Ng1
L2
Nc2
Ng2
Ec
L3
Nc3
Ng3
L4
Nc4
Ng4
N2
N2
N1
Vg
Q1
Q2
Rb2
Rb1
RL1
RL2
N3

Fig. 6 (left.) Polarity reversible magnetic-transistor amplifier, all values same as the equivalent shown in Fig. 1

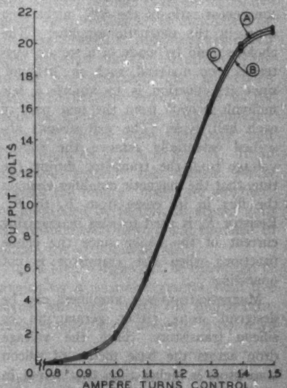


Fig. 7 (right). Effect of supply voltage variation

A—120-volt supply
B—108-volt supply
C—132-volt supply

so that the bias current equals the maximum average load current divided by the gain of the transistor.

$$R_b = \frac{E_c}{I_b} = \frac{E_c\beta}{I_L}$$

where

E_c = average transistor supply voltage
I_b = average bias current
β = current gain of the transistor
I_L = average load current (maximum)

For cases where high sensitivity is required, the gain of the magnetic amplifier portion of the circuit can be improved using any of the more sensitive core materials. Fig. 3, curve B shows the transfer characteristic of a unit using an 80% nickel-iron core material and the circuit if Fig. 1. Since the squareness ratio of the 80% nickel-iron is lower, there is an appreciable output at zero control current. This is to be expected since the transistor is biased "on" in absence of output from the magnetic amplifier. Self-saturating magnetic amplifiers made with cores having a poor squareness ratio will exhibit reduced output because of the voltage absorbed in changing flux from B_r to B_m. A more desirable transfer characteristic can be obtained as shown by curve C in Fig. 3 by the simple expedient of modifying the waveshape of the transistor supply voltage.

The circuit diagram of Fig. 4 is the same as Fig. 1 except that a small choke has been added in series with the transistor supply source. The size of this choke will be determined by the squareness ratio of the 80% nickel-iron cores in the magnetic amplifier. The choke should be made of a square hysteresis loop material, such as Orthonol, since its function is to absorb a few millivolt-seconds from the first part of each half cycle. The volt-seconds absorbed essentially remove the supply voltage from the transistor during the time that the magnetic amplifier changes the flux in its cores from B_r to B_m. Resistor R_d is used to pass magnetizing current of the choke since the choke functions when the transistor is not operating.

Magnetic-transistor amplifiers can be designed using either germanium or silicon transistors. Since the voltage drop across the base diodes of silicon transistors is higher than that found in germanium transistors, the magnetic amplifier for use with silicon transistors must have a higher output voltage than one designed for use with germanium transistors. Since the maximum reverse voltage for silicon transistors is considerably higher than for germanium, this presents no problem.

Circuit Modifications

There are additional magnetic-transistor circuits using the same scheme of that shown in Fig. 1 which will also exhibit self-regulation. The circuit shown in Fig. 5 may be used to deliver an a-c output. The transistor in this circuit is connected across the d-c terminals of a full-wave rectifier bridge which is in series with the load and a-c supply source. When the transistor switch is open-circuited, the bridge presents a high impedance to the load and when short-circuited offers a low impedance.

The transistor is supplied with a bias to turn it on and the magnetic amplifier is used to gate the transistor switch as was explained previously.

Fig. 6 shows the circuit of a push-pull magnetic-transistor amplifier. This circuit is made up of two single-ended units connected so that a given direction of signal increases the output of one of the single-ended units and decreases the output of the other. This circuit has d-c output and can be taken differentially to obtain reversible polarity d-c output. The circuit also exhibits self-regulation, having the same properties of that shown in Fig. 1.

Experimental Results

Tests were conducted on magnetic-transistor amplifiers using both silicon and germanium transistors with 50% and 80% nickel-alloy core materials. Fig. 7 shows a typical transfer characteristic of the circuit in Fig. 1 as well as the effects on ± 10% variation in supply voltage. It is of interest that regulation can be maintained past the knee in the region of maximum output. This is accomplished by designing the magnetic amplifier such that at full output and minimum supply voltage the core flux is operated from knee to knee. This flux excursion is the maximum volt-seconds of the core and is a constant; hence, supply voltages in excess of this minimum do not adversely affect the output.

The regulation of the circuit with variations in load is excellent when germanium transistors are used since the saturation resistance of these transistors is one ohm or less. The regulation with load using silicon transistors is higher since the saturation resistance of silicon transistors is much higher than germanium. The variation in output was small over temperature ranges from −45 to +52 degrees centigrade. The effect of temperature change was to translate the transfer curve by an amount predicted by the change in hysteresis loop width as a function of temperature, approximately 0.07% per degree centigrade.

A check of the speed of response showed that with low N_c^2/R_c ratios, the speed of response was less than 1 cycle. It is interesting that the circuit exhibits essentially the same response for both rise and fall times. This probably can be attributed to the limiting of circulating currents flowing in the postfiring interval.

Fig. 8 (A) shows the output of the magnetic amplifier alone, when terminated into a resistive load. The rise time of the voltage which occurs at the firing angle requires 200 microseconds. Fig. 8 (B) shows the output of the magnetic-transistor amplifier under the same conditions. The switching time of the transistor is 24 microseconds. This is the maximum amount of time the transistor is in its class A region during each half-cycle. The fact that the transistor switches very much faster than the magnetic amplifier output in this combination device permits operation of this circuitry at much higher frequencies than are possible with only the magnetic amplifier.

The transistor functions more nearly as a perfect switch than does the magnetic amplifier, and is less frequency sensitive than are the windings of the saturated cores. Therefore, the output of the magnetic-transistor amplifier is changed

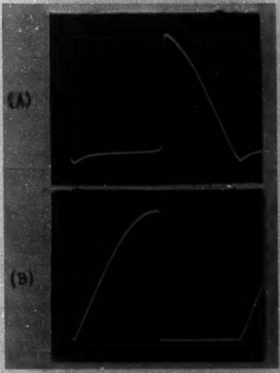


Fig. 8. Comparison of waveforms of magnetic amplifier alone (A); and output of magnetic-transistor amplifier (B)

less by variations in supply frequency than the magnetic amplifier alone.

Conclusions

Magnetic amplifiers and transistors can be combined in a hybrid amplifier to produce a variety of advantages, foremost of which is self-regulation for variations in supply voltage.

In addition to this self-regulation for variations in supply voltage, the circuitry provides high gain, frequency and temperature stability, and excellent load characteristics.

The design of the magnetic-transistor amplifier is simple since the magnetic amplifier of the circuit is a conventional self-saturating circuit.

The use of this circuitry has the additional advantage of being suitable for use at higher supply frequencies since the transistor is switched at approximately 10 times the speed the magnetic amplifier output normally is switched.

References

1. THE TRANSACTOR: A SELF-SATURATED TRANSFORMER, PART I, Allen S. Rosenstein. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, May 1958, pp. 120–42.

2. SHUNT-COUPLED MAGNETIC AMPLIFIER CIRCUITS, R. M. Hubbard. *Ibid.*, vol. 78, May 1959, pp. 124–31.

Magnetic Amplifier Binary-to-Analog Conversion

I. DANYLCHUK
ASSOCIATE MEMBER AIEE

D. KATZ
ASSOCIATE MEMBER AIEE

THE PROBLEM of digital to analog conversion occurs frequently in systems using digital computers. Where power gain is needed, where a high degree of accuracy must be maintained, and where speed of response is not a limiting factor, magnetic amplifiers have special utility in decoding of digital information. Based on the saturable reactor with multiple control windings proportioned in a binary scale, a magnetic amplifier circuit has been devised to convert a 10-digit parallel binary-coded input to an analog voltage ranging from 0 to 102.3 volts average with an accuracy of better than one part in a thousand. With the 10-digit input, three saturable reactors are used for the basic conversion to analog and a subsequent stage of self-balancing magnetic amplifier provides additional gain to drive a servomagnetic amplifier. In the over-all conversion circuit, the inherent feedback, signal mixing, and power gain capabilities of magnetic amplifiers appear to result in a simpler and more reliable circuit than could be achieved with other techniques.[1] This paper discusses the application of magnetic amplifiers in a digital to analog conversion.

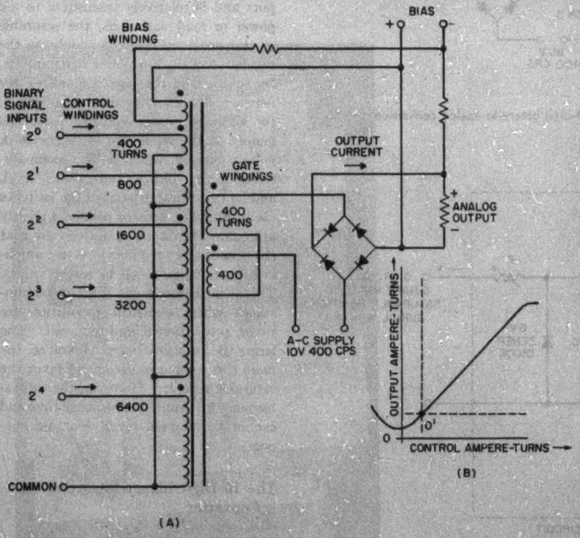


Fig. 1(A). Saturable reactor for 5-digit binary-to-analog conversion. (B). Shift of coordinate axes of transfer characteristic

The Saturable Reactor as a Basic Converter

Binary to analog conversion can be considered as a problem in signal weighting and addition. In the binary system only two symbols are used, 0 and 1. These symbols represent the coefficients of integral powers of the base 2. For example, the binary number 01011 represents $0\times 2^4+1\times 2^3+0\times 2^2+1\times 2^1+1\times 2^1$, or the decimal number 11. To convert a binary number to its analog equivalent, each coefficient is multiplied by the base 2 raised to the power associated with the digit and the quantities so obtained are summed. The conversion factors of the five binary digits in the example, going from the least significant digit (2^0) to the most significant digit (2^4), would be 1, 2, 4, 8, and 16. In an electrical system, the signals representing the binary "ones" are weighted, or multiplied by factors, in accord with these conversion factors and summed to obtain the analog equivalent.

Five binary digits can be converted to analog by a saturable reactor with five signal windings. With the simple saturable reactor, good linearity is obtained

---

Paper 59-1044, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-IRE Nonliner Magnetics and Magnetic Amplifiers Conference, Washington, D. C., September 23–25, 1959. Manuscript submitted June 23, 1959; made available for printing July 16, 1959.

I. DANYLCHUK and D. KATZ are with Bell Telephone Laboratories, Inc., Whippany, N. J.

and the effects of power supply and load variations are minimized.[3] Fig. 1(A) shows a saturable reactor circuit with design values used in the conversion of 5 binary digits. It is a series-connected saturable reactor with gate windings on 4-79 permalloy tape wound cores. The binary input signals are represented by the presence of a 1-milliampere current in the control winding for a binary "1" and the absence of current for a binary "0." As indicated in the figure, the number of turns in each control winding is weighted according to the significance of the binary digit with which it is associated. Since the average gate or output current is directly proportional to the sum of the control ampere-turns in the saturable reactor, the output current is the analog equivalent of the binary input. To maintain high conversion accuracy in the 10-digit converter described later, the bias arrangement shown in Fig. 1 is used. It biases the saturable reactor beyond the exciting current portion of its transfer characteristic into the linear range of operation. The bias shifts the co-ordinate axes as shown by the dashed lines in Fig. 1(B).

The basic feature of the saturable reactor as a converter is that it proportions and adds the several binary inputs at a low voltage and power level and gives the result at a higher output power. Since its output is a direct function of the inputs and is relatively insensitive to a-c power or load variations, the saturable reactor accuracy depends mainly on the regulation of the low-power signal inputs. By contrast, in the conventional resistor networks for binary to analog conversion, the accuracy of the conversion depends mainly on the regulation of the network voltage supply, which is at a comparatively high voltage and power level, and on the characteristics of the switches used. Further, resistor networks greatly attenuate the input quantities so that the output often requires large d-c amplification to bring it up to useful values. This comparison holds for any circumstance where electrical quantities are to be proportioned and summed. The larger the number of signals involved, the more this comparison seems to favor the saturable reactor. However, the saturable reactor is limited in response time and cannot be adapted readily for a-c outputs.

The 10-Digit Binary-to-Analog Converter

The saturable reactor is applied in the conversion of 10-digit binary signals from

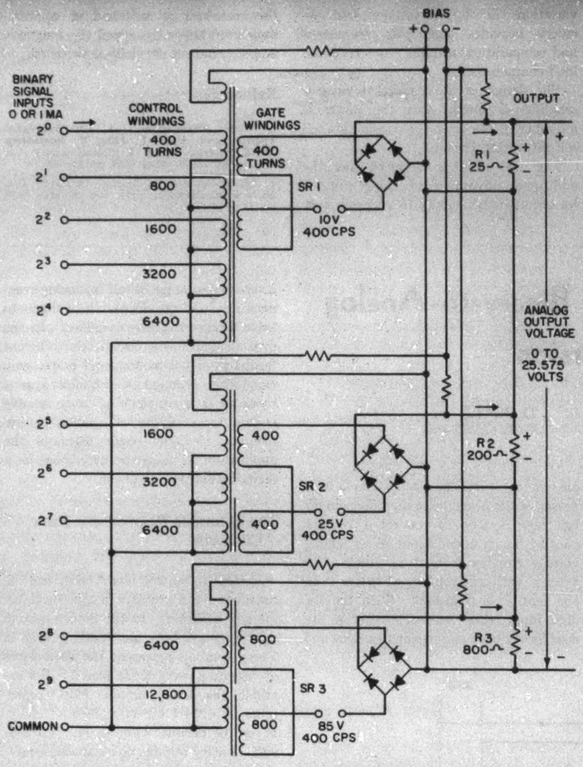

BIAS
+
−
BINARY SIGNAL INPUTS 0 OR 1 MA
CONTROL WINDINGS
GATE WINDINGS
OUTPUT
2^0
400 TURNS
400 TURNS
R1 25Ω
+
−
+
2^1
800
SR 1
2^2
1600
10V 400 CPS
2^3
3200
2^4
6400
ANALOG OUTPUT VOLTAGE
0 TO 25.575 VOLTS
2^5
1600
400
R2 200Ω
2^6
3200
SR 2
2^7
6400
400
25V 400 CPS
2^8
6400
800
R3 800Ω
2^9
12,800
SR 3
COMMON
800
85 V 400 CPS

Fig. 2. Three saturable reactors for 10-digit binary-to-analog conversion

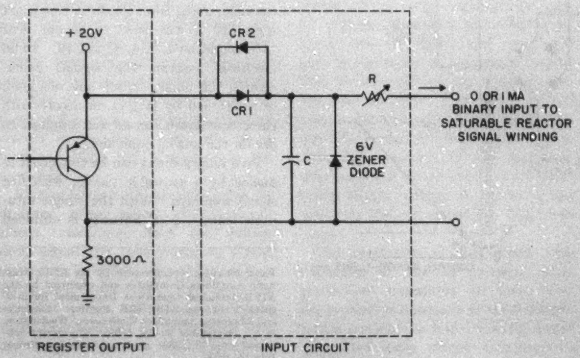

+20V
CR2
CR1
R
0 OR 1 MA
BINARY INPUT TO SATURABLE REACTOR SIGNAL WINDING
C
6V ZENER DIODE
3000Ω
REGISTER OUTPUT
INPUT CIRCUIT

Fig. 3. Input circuit to saturable reactors

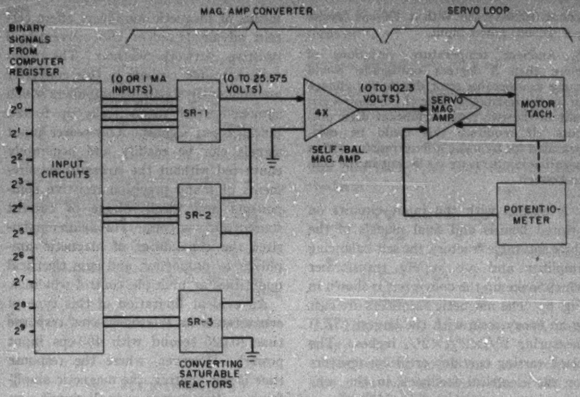


Fig. 4. Block diagram of converter and servoloop

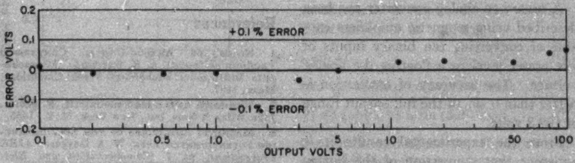


Fig. 5. Error voltage versus output voltage

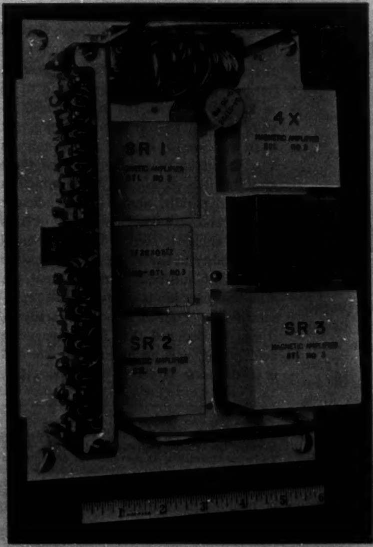


Fig. 6. Model of magnetic amplifier converter

a computer to an analog voltage which must have enough power to drive a servomagnetic amplifier. The 5-digit saturable reactor of Fig. 1 is supplemented by two other saturable reactors to make a 10-digit converter. As is shown in Fig. 2, the three saturable reactors are arranged so that their output voltages are added to obtain the analog of the 10-digit binary input. Additional saturable reactors are used because in this case it is not practical to put all ten control windings on one unit. Continuing the signal weighting in control windings for the more significant digits would require as much as 2^9 times 400 or 204,800 turns for the most significant digit. For this reason, partial signal weighting in *SR2* and *SR3* is accomplished by their respective load resistors, *R2* and *R3*. In the circuit of Fig. 2, *R2* is eight times the resistance of *R1*, so that the output voltage of *SR2* per ampere-turn of input is eight times that of *SR1*. Therefore, the control turns required on *SR2* to continue the binary scale are reduced by a factor of eight. In *SR3*, there is a change in gate winding turns and output resistance so that the control turns are reduced by a factor of 16. The a-c voltages supplied to the saturable reactors increase with output resistance to operate each one in the linear region.

For the accurate performance of this converter it is necessary that the signal inputs be invariant. In this application the 10-digit binary information comes from a computer register. The output of the register across a transistor switch is +0.2 volts for the logical "0" and +20 volts in series with a 3,000-ohm resistor for logical "1." The 20-volt supply to the transistor switches varies ±0.25%. This variation would result in an analog conversion error greater than a desired maximum of 0.1%. The input circuit of Fig. 3 uses a 6-volt Zener diode to provide a constant voltage signal. The two rectifier diodes and a capacitor eliminate errors due to the +0.2 volt for logical "0" and rectification of induced second harmonic voltages from the saturable reactors. A variable resistor (*R*) adjusts the nominal 1-milliampere signal current to compensate for variations in winding turns and Zener voltage from unit to unit.

The analog output of the saturable reactors shown in Fig. 2 is 0 to 25.575 volts average. A final stage of the converter, a self-balancing magnetic amplifier[3], acts as a buffer stage with very high input impedance, large power gain, and good linearity. This magnetic amplifier, having a voltage gain of four, brings the analog output up to 102.3 volts average which is needed to drive to servomagnetic amplifier. The servoloop positions a stylus which plots the analog value of the computer information. A block diagram of the circuit from computer input to servomotor is shown in Fig. 4.

Operation of the magnetic amplifier converter, consisting of the three saturable reactors and the self-balancing magnetic amplifier in Fig. 4, has been satisfactory over the full range of output voltage (0 to 102.3 volts d-c). Data taken with models under actual operating conditions show that the accuracy requirement of 0.1% of full converter output has been met. Fig. 5 gives the measured errors of a typical converter over the full output range. The general electrical characteristics of the magnetic amplifier converter can be summarized as follows:

1. The d-c signal power from the input circuit to the converting magnetic amplifiers for 1 digit is about 10 milliwatts.
2. The converter is capable of delivering 2 watts at 102.3 volts without any loss in accuracy.
3. The a-c power input to the converter is less than 20 watts.
4. The frequency and the voltage variation of ±5% have negligible effect on the required accuracy of 0.1% of full output.
5. The over-all response of the converter is about 0.025 second with a 400-cps (cycles per second) power input.
6. Ambient temperature variations of more than 5 degrees centigrade would cause a noticeable error due to the change in the d-c resistance value of the control circuit in the two most significant digit outputs. If required, this could be compensated for by using a fixed resistor with a negative temperature coefficient in the control circuit.

A chassis with the input circuits on printed boards and final models of the three saturable reactors, the self-balancing amplifier and a-c supply transformer which make up the converter, is shown in Fig. 6. The magnetic amplifiers are cast in an epoxy-resin with the largest (*SR3*) measuring $2\frac{5}{8} \times 2\frac{5}{8} \times 2\frac{3}{4}$ inches. The black casting contains precision resistors for the electrical feedback in the self-balancing magnetic amplifier (*4X*).

Conclusions

A binary-to-analog converter has been described using magnetic amplifiers capable of converting ten binary inputs of low power level to a floating d-c analog voltage. The accuracy of conversion is better than 0.1% of the full output range of 102.3 volts.

From the experimental results it is apparent that conversion of the binary coded information to analog voltages by means of magnetic amplifiers offers certain advantages over the conventional resistive network method. The magnetic amplifier circuits in this converter are inherently stable d-c amplifiers which sum and weigh magnetically up to ten binary input signals. Low-power input signals can be readily and accurately converted without the stringent requirements of many precision resistive components and ideal voltage or current sources and switches. This again emphasizes the capabilities of magnetic amplifiers to proportion and sum electrical quantities in multiple control windings.

An evident limitation of this type of converter is its relatively slow response time (0.025 second with 400-cps input power). However, where the response time is not limiting, the magnetic amplifier can often permit over-all circuit simplification and is an attractive component in military systems because of its ruggedness and reliability.

References

1. NOTES ON ANALOG-DIGITAL CONVERSION TECHNIQUES (BOOK), A. K. Susskind. Massachusetts Institute of Technology Press, Cambridge, Mass., 1957.

2. MAGNETIC AMPLIFIERS (BOOK), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955.

3. MAGNETIC AMPLIFIERS OF THE SELF-BALANCING POTENTIOMETER TYPE, W. A. Geyger. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 71, 1952 (Jan. 1953 section), pp. 383–95.

Vol,
Pt. 1
195

The A-C Resistance of Solid Magnetic Wires

M. KAMAL GOHAR
NONMEMBER AIEE

WITH THE development of new types of ferromagnetic materials, magnetic conductors are increasingly used in connection with automatic control and electric power, and for industrial and scientific purposes. Here, the a-c resistance varies not only with the frequency but also with the current, as the permeability is not constant over the cross section of the wire. Previous authors[1–3] have given experimental results, and have used qualitative solutions. To make the solution possible, the theoretical analysis has been mainly based on assuming average constant permeability over the cross section of the conductor; this, however, is not actually the case, as the permeability varies from the axis to the circumference; see Figs. 1 through 3. As a consequence, previous authors have not found it possible to obtain numerical correlation between the magnetic properties of such wire deduced from its impedance measurement and those measured by direct magnetic methods. There is still less possibility of predicting the a-c resistance or inductance of these wires from their measured magnetic properties.

To overcome this difficulty, a solution is presented here based, on the new representation of the **B-H** curve previously suggested by the writer,[4] and which can be applied with advantage. It will be used to obtain the a-c resistance of solid magnetic wires from their measured magnetic properties. The theoretical results, verified experimentally, are found to be in accordance with the experimental results obtained from previous authors.

In this case the **B-H** curves as shown in Fig. 4 can be represented by the three straight lines *oa*, *ab*, and *bc*, in which *oa* and *bc* are parallel, and *ab* is perpendicular to the **H**-axis. The simplicity of this method is clearly seen as the incremental permeability in the two regions *oa*; and *bc* is constant and equal, which is not the case for other kinds of **B-H** curve representations. Following the same analysis, it can be deduced that if the magnetic field in Fig. 4 is an alternating one and is in the form **H** sin ωt where $|\mathbf{H}|<|\mathbf{H}_0|$,

---

Paper 59-1086 recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted July 22, 1958; made available for printing June 25, 1959.

M. KAMAL GOHAR is with Cairo University, Cairo, Egypt.

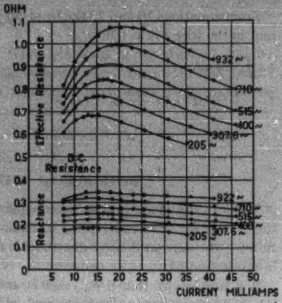


Fig. 1. A-c resistance of a magnetic wire

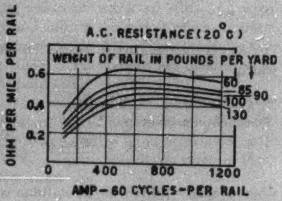


Fig. 2. A-c resistance of a steel rail

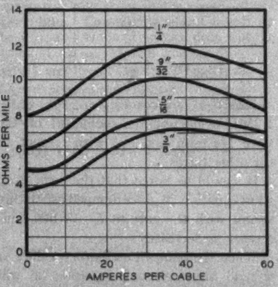


Fig. 3. A-c resistance of steel ground wire, grade EB7 strand cable, 60 cycles

the corresponding flux density **B** will be

$$\mathbf{B}=\mu\mathbf{H}\sin\omega t \tag{1}$$

where μ is the slope of oa or bc. While, if the equation of the straight line bc in Fig. 4 is

$$\mathbf{B}=\mu\mathbf{H}+C_0 \tag{2}$$

and the alternating field in equation 1 has its maximum value $|\mathbf{H}|>|\mathbf{H}_0|$, then the corresponding flux density will be

$$\mathbf{B}=\mu\mathbf{H}\sin\omega t+C_0(\omega t) \tag{3}$$

where $C_0\ (\omega t)$ is a Fourier series in the form of

$$a_0+\Sigma a_n\cos n\omega t+\Sigma b_n\sin n\ \omega t \tag{4}$$

representing the constant value C_0 when the magnitude of the field is greater than $|\mathbf{H}_0|$.

According to Fig. 4, $C_0(\omega t)$ will contain odd harmonics only and will take the form

$$C_0(\omega t)=\frac{2C_0}{\pi}\sum\frac{\cos n\,\theta_1-\cos n\,\theta_2}{n}\times \sin n\,wt\quad n\text{ odd}$$

$$=\frac{4C}{\pi}\sum\frac{\cos n\theta_1}{n}\sin n\ \omega t\quad n\text{ odd} \tag{5}$$

In this case $a_0=a_n=0$, $\theta_2=\pi-\theta_1$, and $\mathbf{H}\sin\theta_1=C_0$

It should be noted that the problem here deals only with the value of the a-c resistance and that hence there is no need to consider the harmonic terms in the calculations (other than the fundamental one) if the values of θ_1 and θ_2 are satisfactorily determined. For example, if the current is in the form

$$i_1\sin\omega t \tag{6}$$

and the applied voltage becomes

$$V_1\sin(\omega t+\phi_1)+v_3\sin(3\omega t+\phi_3)+\ldots \tag{7}$$

the power, and hence the effective resistance, will depend only on the magnitudes of the current and voltage of the fundamental frequency only, since the average power will be

$$(v_1 i_1/2)\cos\phi_1 \tag{8}$$

As may be deduced from the following analysis, the harmonic terms are not of value except when it is required to determine the values of θ_1 and θ_2 after the second approximation.

Note that in the theoretical analysis the asymptotic expansion series for the high values of Bessel functions will be adopted; for example when $z>8$,

$$\text{ber}\ z+j\ \text{bei}=\frac{0.399}{\sqrt{z}}\epsilon^{\frac{z}{\sqrt{2}}}\epsilon^{j\frac{z}{\sqrt{2}}+\frac{\pi}{8}}$$

$$=|M(z)|\epsilon^{j\alpha_z}=M_0(z) \tag{9}$$

$$\text{bei}'\ z-j\ \text{ber}'\ z=\frac{0.399}{\sqrt{z}}\epsilon^{\frac{z}{\sqrt{2}}}\epsilon^{j\frac{z}{\sqrt{2}}-\frac{3\pi}{8}}$$

$$=|M(z)|\epsilon^{j\alpha_z'}=M'(z) \tag{10}$$

For smaller values of z, the other forms of Bessel functions are to be similarly treated but they will be of less importance.[5]

Theoretical Analysis

If an alternating current in the form

$$I\sin\omega t \tag{11}$$

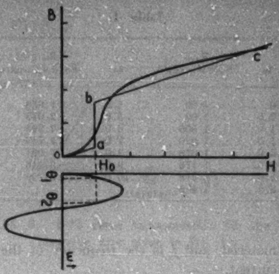


Fig. 4. B-H curve representation

passes in a magnetic conductor of radius a and whose **B-H** curve is represented by $oabc$ as shown in Fig. 4, the maximum magnetic field

$$|\mathbf{H}_a|=\left|\frac{2I}{a}\right| \tag{12}$$

will be at the circumference of the wire. Two cases must now be considered:

Case 1

In case 1, if $|\mathbf{H}_a|<|\mathbf{H}_0|$, the permeability μ can satisfactorily be considered constant over the cross section of the wire. The current density i is given in equation 13 which has been proved by many physicists and electrical engineers.[6]

$$i=A_0(\text{ber}\ mr+j\ \text{bei}\ mr)+\mathbf{B}_0(\text{ker}\ mr+j\ \text{kei}\ mr)$$

$$=A_0M_0(mr)+\mathbf{B}_0Y_0(mr) \tag{13}$$

where, for a solid wire $\mathbf{B}_0=0$

$$A_0=\left(\frac{Im}{2\pi a}\right)/M'(ma),\ m^2=\frac{8\pi^2 f\mu}{\rho}$$

where ρ is the resistivity of the magnetic

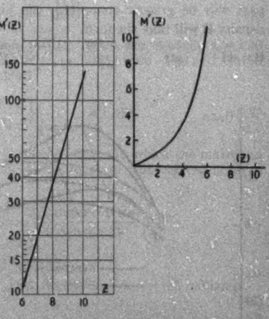


Fig. 5. M′(z)–z chart

Table I

z_a	$\frac{R_{a\text{-}c}}{R_{d\text{-}c}}$ μ constant	z_a	$\frac{R_{a\text{-}c}}{R_{d\text{-}c}}$ μ constant
0	1.00	4.0	1.678
0.5	1.00	5.0	2.043
1.0	1.005	6.0	2.394
1.5	1.026	7.0	2.743
2.0	1.078	8.0	3.094
2.5	1.175	9.0	3.446
3.0	1.318	10.0	3.799
3.5	1.492		

material, and f is the frequency of the current.

The magnetic field **H** at any radius r will be

$$H=\frac{2\int_0^r 2\pi r i dr}{r}=\frac{2I}{a}\,\frac{M'(z)}{M'(z_a)} \tag{14}$$

where $z=mr$ and $z_a=ma$.

Fig. 5 represents the values of $M'(z)$ for different values of z, from $z=0$ to $z=10$. The ratio of the a-c resistance to the d-c one, in this case $(R_{a\text{-}c}/R_{d\text{-}c})_{\mu\text{constant}}$, was found by previous authors[6] to be equal to the real part of

$$\frac{z_a}{2}\,\frac{M_0(z_a)}{M'(z_a)}$$

$$=\frac{\text{ber}\, z_a\, \text{bei}'\, z_a-\text{bei}\, z_a\, \text{ber}'\, z_a}{\text{ber}'^2\, z_a+\text{bei}'^2\, z_a} \tag{15}$$

These values are shown in Table I from $z=0$ to $z=10$.

Case 2

In case 2, the cross section of the wire will be divided into two parts: the first one is from $r=0$ to $r=r_0$ at which the maximum magnetic field will be equal to $\mathbf{H}_0$. The current density distribution will take its original ordinary form as given in equation 13:

$$i=A_0'M_0(z) \tag{16}$$

The second part will be from $r=r_0$ to $r=a$. As the maximum magnetic field in this part will be greater than $|\mathbf{H}_0|$, the flux density **B** will take the form

$$\mathbf{B}=\mu\mathbf{H}+C_0'(\omega t) \tag{17}$$

where $C_0'(\omega t)$ is the Fourier series representing the constant value C_0 of the flux density **B** in equation 2. $C_0'(\omega t)$ is a function of $\cos\theta_1$, r, or z. According to Fig. 4 and equation 5, it should have a zero value when $r=r_0$ or $z=z_0$.

As illustrated in Appendix I, a convenient representation for the coefficient of the fundamental frequency term in $C_0'(\omega t)$ is

$$\left(\frac{4C_0}{\pi\mu}\right)\tanh\alpha\,(z-z_0)\epsilon^{j\frac{z-z_a}{\sqrt{2}}} \tag{18}$$

where $\alpha=0.7+\epsilon^{-z_0}$

and the differential equation for the magnetic field distribution **H** is found to be

$$\frac{d^2\mathbf{H}}{dz^2}+\frac{1}{z}\,\frac{d\mathbf{H}}{dz}=\frac{\mathbf{H}}{z^2}-j\mathbf{H}$$
$$=j\left(\frac{4C_0}{\pi\mu}\right)\tanh\alpha(z-z_0)\epsilon^{j\frac{z-z_a}{\sqrt{2}}} \tag{19}$$

Considering large values of z and neglecting the coefficients of small terms, equation 20 is a satisfactory solution for differential equation 19.

$$\mathbf{H}=\frac{4\pi}{m}A_0'M'(z)+\frac{4\pi}{m}\,\mathbf{B}_0'Y'(z)-$$
$$(0.8+j0.4)\frac{4C_0}{\pi\mu}\tanh\alpha(z-z_0)\epsilon^{j\frac{z-z_a}{\sqrt{2}}} \tag{20}$$

Using the relation between i and **H**, which is

$$\frac{d(r\mathbf{H})}{dr}=4\pi r i \tag{21}$$

the current density distribution i can be deduced from equation 20, which gives

$$i=A_0'M_0(z)+B_0'Y_0(z)-\frac{(0.8+j0.4)C_0m}{\pi^2\mu}\times$$
$$\left[\alpha\,\text{sech}^2\,\alpha(z-z_0)+\frac{j}{\sqrt{2}}\tanh\alpha(z-z_0)+\right.$$
$$\left.\frac{1}{z}\tanh\alpha(z-z_0)\right]\epsilon^{j\frac{z-z_a}{\sqrt{2}}} \tag{22}$$

when $z=z_0$, equations 16 and 22 must be equal; hence

$$\mathbf{B}_0'=\frac{(0.8+j0.4)C_0m}{Y_0(z_0)\pi^2\mu}\,\alpha\epsilon^{j\frac{z_0-z_a}{\sqrt{2}}} \tag{23}$$

From equations 16 and 22 the total current in the wire I is as follows:

$$I=\int_0^a 2\pi r i dr$$

$$I=\frac{2\pi a}{M}\,A_0'M'(z_a)+\frac{2\pi a}{m}\,\mathbf{B}_0'Y'(z_a)-$$
$$\frac{2\pi r_0}{m}\,\mathbf{B}_0'y'(z_0)-\frac{4C_0}{\pi\mu}\,\frac{(04+j0.2)}{m}\times$$
$$z_a\tanh\alpha(z_a-z_0) \tag{24}$$

If the values of A_0' and $\mathbf{B}_0'$ from equations 24 and 23 are substituted in equation 22, the current density distribution i in the second part can be deduced.

It is well known that if $R_{a\text{-}c}$ and L_i are the a-c resistance and the internal inductance of the wire respectively, then

$$R_{a\text{-}c}+j\omega L_i=\frac{i_a\rho}{I} \tag{25}$$

i.e., $R_{a\text{-}c}$ is the real part in $(i_a\rho)/I$ from which

$$\frac{(R_{a\text{-}c})}{R_{d\text{-}c}}=\frac{(R_{a\text{-}c})}{R_{d\text{-}c\,\mu\,\text{constant}}}+\frac{\phi(z_a,\, z_0)}{I}+$$
$$\frac{\Psi(z_a,\, z_0)}{I} \tag{26}$$

where $\phi(z_a,z_0)=[(0.178\; C_0)/\mu]\; az_a[\tanh\alpha(z_a-z_0)-1.43\;\alpha\;\text{sech}^2\,\alpha(z_a-z_0)+1.43\alpha\sqrt{\frac{Z_0}{Z_a}}\epsilon^{\frac{z_0-z_a}{\sqrt{2}}}\cos\sqrt{2}(z_0-z_a)-0.715\alpha\sqrt{\frac{Z_0}{Z_a}}\epsilon^{\frac{z_0-z_a}{\sqrt{2}}}\sin\sqrt{2}(z_0-z_a)\Bigg]$ and $\Psi(z_a,z_0)=-0.286\frac{c_0\alpha a z_a}{\mu}\Bigg[\sqrt{\frac{Z_0}{Z_a}}\epsilon^{\frac{z_0-z_a}{\sqrt{2}}}\cos(\sqrt{2}z_0+1.256)-z_0/z_a\;\cos(3z_0/\sqrt{2}-z_a/\sqrt{2}+1.256)]$ neglecting the coefficient of the term $1/z\tanh\alpha(z_a-z_0)$ with respect to the other terms.

It should be noted that the term $\Psi(z_a,z_0)$ in equation 26 gives damped oscillating values with the current I. (When I increases, z_0 decreases.)

When the damped oscillatory values of $\Psi(z_a,z_0)$ are neglected, equation 26 will reduce to

$$\frac{R_{a\text{-}c}}{R_{d\text{-}c}}=\left(\frac{R_{a\text{-}c}}{R_{d\text{-}c}}\right)_{\mu\text{ constant}}+\frac{\phi(z_a,\, z_0)}{\sqrt{2}I_{\text{rms}}} \tag{27}$$

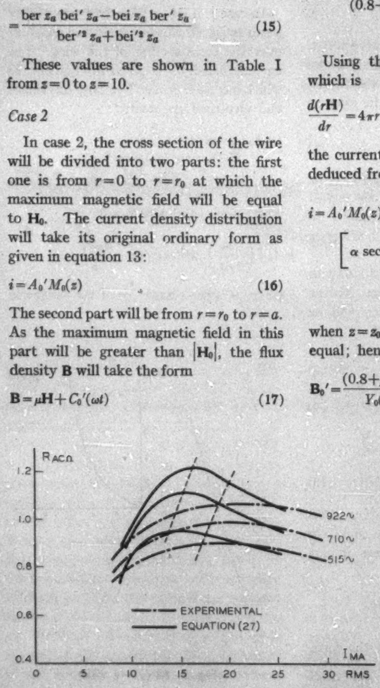


Fig. 6 (left). Comparison between experimental and theoretical results

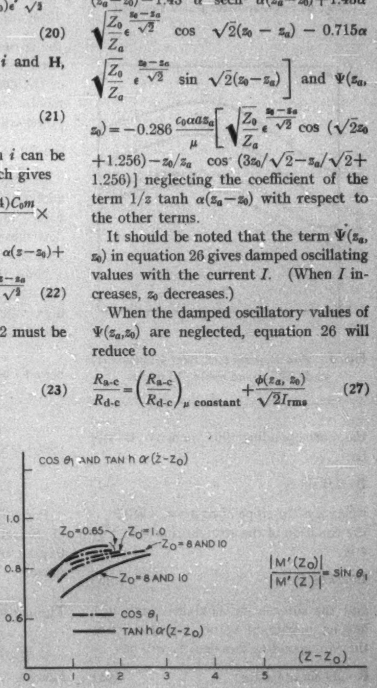


Fig. 7 (right). Comparison between $\cos\theta_1$ and $\tanh\alpha(z-z_0)$

where $I=\sqrt{2}\,I_{rms}$

The first term in equation 27 can be taken from Table I or from the textbooks.[6] The second term in equation 27 is given in equation 26.

From equation 27 it may be seen that

$$\frac{R_{a\text{-}c}}{R_{d\text{-}c}}=\left(\frac{R_{a\text{-}c}}{R_{d\text{-}c}}\right)_{\mu\ \text{constant}}$$

when $z_a=z_0$; and if $z_a>z_0$, the resistance ratio given in equation 27, will increase at first with the current until the increase in the current begins to be greater than the increase in $\phi(z_a,z_0)$. The resistance ratio will then begin to decrease with the current I. Fig. 6 shows a comparison between the experimental a-c resistance extracted from Harrison, et al.,[1] and the theoretical values obtained from equation 27. The procedure and method of calculations are illustrated in Appendix II.

The variation of the a-c resistance as obtained from equation 27 and shown in Fig. 6 is found to be in accordance with the experimental results as given in Figs. 1, 2, and 3.

Appendix I

The alternating voltage drop must be constant at any point in the cross section of the wire. Considering a unit length of the wire, two kinds of voltage drop will be produced. The resistive voltage drop $=i\rho$ and the reactive voltage drop $=j\omega\int_r^a B\,dr+$ the constant external reactive voltage drop.

According to equations 5 and 17, $\mathbf{B}=\mu\mathbf{H}+C_0'(\omega t)$; hence

$$i\rho+j\omega\int_r^a \mu H\,dr+j\omega\int_r^a C_0'(\omega t)\,dr = \text{constant} \tag{28}$$

Differentiating equation 28 with respect to r yields

$$\frac{di}{dr}-j\omega\mu\mathbf{H}-j\omega C_0'(\omega t)=0 \tag{29}$$

The magnetic field **H** at any point in the second part of the wire is given by

$$\mathbf{H}=\frac{r_0\mathbf{H}_0}{r}+\frac{2\int_{r_0}^{r}2\pi ri\,dr}{r} \tag{30}$$

$$\text{i.e.,}\ \frac{d(r\mathbf{H})}{dr}=4\pi ri=r\frac{d\mathbf{H}}{dr}+\mathbf{H} \tag{31}$$

From equation 31

$$\frac{di}{dr}=\frac{1}{4\pi}\left(\frac{d^2\mathbf{H}}{dr^2}+\frac{1}{r}\frac{d\mathbf{H}}{dr}-\frac{\mathbf{H}}{r^2}\right) \tag{32}$$

When the value of di/dr from equation 32 is substituted into equation 29, the differential equation for the magnetic field distribution **H** in the second part of the wire will be

$$\frac{d^2\mathbf{H}}{dr^2}+\frac{1}{r}\frac{d\mathbf{H}}{dr}-\frac{\mathbf{H}}{r^2}-j\frac{4\pi\omega\mu\mathbf{H}}{\rho}=j\frac{4\pi\omega}{\rho}C_0'(\omega t) \tag{33}$$

Letting $4\pi\mu\omega/\rho=m^2$, and $mr=z$, yields for equation 33

$$\frac{d^2\mathbf{H}}{dz^2}+\frac{1}{z}\frac{d\mathbf{H}}{dz}-\frac{\mathbf{H}}{z^2}-jH=j\frac{C_0'(\omega t)}{\mu} \tag{34}$$

considering the fundamental frequency only; then from Fig. 4, and equations 5 and 17, $C_0'(\omega t)$ in equation 34 must have a direction, as that of the magnetic field **H**, and a value $=(2C_0/\pi)(\cos\theta_1-\cos\theta_2)_r$, where $(\cos\theta_1-\cos\theta_2)_r$ is a function of r or z representing the value of $(\cos\theta_1-\cos\theta_2)$ at a point at a distance r from the center of the wire.

As a first approximation, C_0/μ is neglected in equation 34. According to equations 13 and 14, the solution for **H** and i will be

$$i=\frac{(Im)}{2\pi a}\,\frac{\text{ber}\,z+j\,\text{bei}\,z}{\text{bei}'\,z_a-j\,\text{ber}'\,z_a}=\frac{(Im)}{2\pi a}\,\frac{M_0(z)}{M'(z_a)} \tag{35}$$

$$\mathbf{H}=\frac{(2I)}{a}\,\frac{\text{bei}'\,z-j\,\text{ber}'\,z}{\text{bei}'\,z_a-j\,\text{ber}'\,z_a}=\frac{(2I)}{a}\,\frac{M'(z)}{M'(z_a)} \tag{36}$$

Using the asymptotic expansion given in equation 10 for $M'z$

$$\mathbf{H}=\frac{2I}{a}\sqrt{\frac{z_a}{z}}\,\epsilon^{\frac{z-z_a}{\sqrt{2}}}\,\epsilon^{j\frac{z-z_a}{\sqrt{2}}} \tag{37}$$

The direction of **H** with respect to the total current I in this case will be given by

$$\epsilon^{j\frac{z-z_a}{\sqrt{2}}} \tag{38}$$

Knowing z_a, I, a, and $\mathbf{H}_0$, one can determine, from equation 36 and Fig. 5, z_0, which is the value of z at which the maximum magnetic field **H** will be equal to $\mathbf{H}_0$. Hence

$$\mathbf{H}_0=\frac{2I}{a}\,\frac{M'(z_0)}{M'(z_a)} \tag{39}$$

From equation 11, if the total current I varies in the form $I\sin\omega t$, then, according to equation 36 the magnetic field will be

$$\frac{2I}{a}\,\frac{M'(z)}{M'(z_a)}\sin\omega t \tag{40}$$

At any point at a distance $r>r_0$ from the center of the wire, or $z>z_0$, the maximum magnetic field $|\mathbf{H}|$ at this point will be greater than $|\mathbf{H}_0|$.

When $z>z_0$

$$\theta_1=\pi-\theta_2$$

and hence

$$(\cos\theta_1-\cos\theta_2)_r=2(\cos\theta_1)_r \tag{41}$$

The value of $(\cos\theta_1)_r$ begins from zero when $z=z_0$, and reaches unity at relatively large values of $(z-z_0)$. At intermediate values of $(z-z_0)$, and from Fig. 4, equations 36 and 39 yield

$$\left|\frac{\mathbf{H}_0}{\mathbf{H}}\right|=(\sin\theta_1)_r=\left|\frac{M'(z_0)}{M'(z)}\right| \tag{42}$$

$$(\cos\theta_1)_r=\sqrt{1-\left[\frac{M'(z_0)}{M'(z)}\right]^2} \tag{43}$$

is plotted in Fig. 7 against $(z-z_0)$ for different values of z_0. In all the cases $(\cos\theta_1)_r$ begins from zero when $z=z_0$, and increases to unity when $(z-z_0)$ is relatively large. But at intermediate values of $(z-z_0)$, $(\cos\theta_1)_r$ is slightly greater for small values of z_0, e.g., when $(z-z_0)=1.0$, $\cos\theta_1=0.91$; when $z_0=0.65$; and if $z_0=8$ or 10, $(\cos\theta_1)=0.85$.

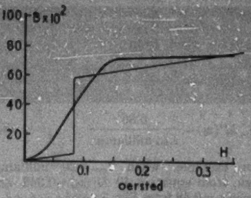


Fig. 8. B-H curve representation for the magnetic wire given in Fig. 1

As seen from Fig. 7, a convenient representation for $(\cos\theta_1)_r$ in this case is found to be

$$(\cos\theta_1)_r=\tanh\alpha(z-z_0) \tag{44}$$

where $\alpha=0.7+\epsilon^{-z_0}$.

From equations 38, 41, and 44 the differential equation will be

$$\frac{d^2\mathbf{H}}{dz^2}+\frac{1}{z}\frac{d\mathbf{H}}{dz}-\frac{\mathbf{H}}{z^2}-j\mathbf{H}=j\frac{4C_0}{\pi\mu}\times \tanh\alpha(z-z_0)\epsilon^{j\frac{z-z_a}{\sqrt{2}}} \tag{45}$$

where $\alpha=0.7+\epsilon^{-z_0}$

Appendix II

The $\mu-\mathbf{H}$ curve at zero frequency for the magnetic material of the wires used by Harrison, et al., may be deduced from Tables III and IV of their paper,[1] from which the corresponding **B-H** curve and its representation are shown in Fig. 8, where $\mathbf{H}_0=0.08$ oersted, $\mu=9{,}000$ and $C_0/\mu=0.55$ ($a=0.02225$ centimeter, the length of wire $=15.25$ centimeters, and its $R_{d\text{-}c}=0.41$ ohm).

From equations 13 and 34

$$z=mr=4.12r\sqrt{f} \tag{46}$$

when $f=515$ cycles per second, $z_a=2.08$ and $M'(z_a)=1.085$ from Fig. 5. From Table I

$$\left(\frac{R_{a\text{-}c}}{R_{d\text{-}c}}\right)_{\mu\text{constant}}=1.078 \tag{47}$$

The value of I_0 rms at which the maximum field will be equal to 0.08 oersted is given by

$$0.08=\frac{2\,I_0\ \text{amperes (rms)}}{10\times0.02225}\sqrt{2}$$

from which

$$I_0\ \text{amp (amperes) (rms)}=6.3\ \text{milliamp} \tag{48}$$

The a-c resistance is a function of the frequency only (a is fixed), and in this case it does not depend on the current until I_0 rms=6.3 milliamp. When I (rms)>6.3 milliamp, then, from equation 39

$$H_0=0.08=\frac{M'(z_0)}{M'(z_a)}\,\frac{0.2\,I\text{ amp (rms)}\sqrt{2}}{0.02225}$$

$$\text{i.e., } M'(z_0)=\frac{6.86}{I_{rms}\text{ milliamp}} \tag{49}$$

For example, when $I_{rms}=20$ milliamp, then, from equation 49 $M'(z_0)=0.343$, and hence $z_0=0.68$ from Fig. 5; and $z_a-z_0=1.4$.

From equation 18

$$\alpha=0.7+0.506=1.206$$

Substituting in equation 27

$$\frac{R_{a\text{-}c}}{R_{d\text{-}c}}=1.078+1.22=2.298$$

$$R_{a\text{-}c}=0.41\times2.298=0.94\text{ ohm}$$

Experimentally, $R_{a\text{-}c}=0.902$ ohm.

References

1. THE ELECTRICAL PROPERTIES OF HIGH PERMEABILITY WIRES CARRYING ALTERNATING CURRENT, E. P. Harrison, G. L. Turney, H. Rowe, H. Gollop. *Proceedings*, Royal Society of London, London, England, Series A, Mathematical and Physical Sciences, vol. CLVII, 1936, pp. 451–79.

2. THE VARIATION IN THE HIGH FREQUENCY RESISTANCE AND PERMEABILITY OF FERROMAGNETIC MATERIALS DUE TO A SUPERIMPOSED MAGNETIC FIELD, J. S. Webb. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 26, 1938, pp. 433–41.

3. CIRCUIT ANALYSIS OF A.C. POWER SYSTEMS (book), Edith Clarke. John Wiley & Sons, Inc., New York, N. Y., 1950, p. 252.

4. A NEW METHOD FOR TREATING NONLINEAR PROBLEMS WITH APPLICATIONS TO IRON-CORED OSCILLATORY CIRCUITS I-SUBHARMONICS DO NOT EXIST, M. K. Gohar. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, 1956 (Jan. 1957 section), pp. 771–76.

5. BESSEL FUNCTIONS FOR ENGINEERS (book), N. W. McLachlan. The Clarendon Press, Oxford, England, 1950, pp. 222–29.

6. ELECTRICAL COILS AND CONDUCTORS (book), H. B. Dwight. McGraw-Hill Book Company, Inc., New York, N. Y., 1945, pp. 155–63.

An Error-Detection System for 5-Unit-Code Teletypewriter Transmission

P. H. BARRY
NONMEMBER AIEE

A. L. WHITMAN
MEMBER AIEE

FOR MANY YEARS teletypewriter users have operated without automatic means for detecting errors. This was principally because a large part of the transmitted material was ordinary English text with a redundancy so large that for the most part there was no difficulty in reconstructing a misspelled word. For example, recognizing that "enginaer" means "engineer" was simple and obvious. For the correct transmission of numerals various expedients were used, such as sending all numerals twice, spelling out the words, as "one," "two," and "five."

More recently, however, two trends have become more marked. The first is toward the transmission of larger proportions of numeric material, roughly designated by the name "data," in contrast to English text, and the second is toward the direct use of transmitted material in automatic computers without its being read and checked by a human operator. This has brought about an increasing demand for an automatic means of detecting when and where transmission errors have occurred so that their possible bad effects may be guarded against.

In view of the wide extent of telegraph and data transmission by means of the common 5-unit telegraph code, there is a need for means of detecting errors in this transmission. This paper describes one such means. The project in its present stage is a trial in actual field situations to determine the service value of the method chosen.

The present project does not include any means for automatic error correction. When an error is found, it is necessary for the attendant at the receiving end to request retransmission of the portion of the material in which the error was detected. In certain automatic switching systems error correction cannot be readily applied. However, its potential advantages are such that it is not rejected at this time but is set aside for future consideration.

Limitation on Detected Errors

Before describing the system, it may be well to consider a basic limitation. No error-detection system (or error-correction system, for that matter) will give satisfactory service if the communication facility generates frequent errors. For each error detected the user has to send a service message and ask for a rerun. In the absence of experience, the frequency of this operation a user would tolerate is a matter of conjecture. (The price for attaining various degrees of accuracy in transmission may be a major controlling factor but is neglected here.) It is expected that the need for 100 reruns in an 8-hour business day would be cause for major complaint; 10 reruns might be acceptable, especially if it were recognized, as it might be in the particular circumstances, that the cost of substantially improved accuracy would be large.

Assuming operation at 100 words per minute, there would be about 300,000 characters transmitted in an 8-hour day. Ten requests for reruns would mean that there had been one error in each 30,000 characters.

The considerations in the foregoing apply to detected errors only; the efficiency of the system with respect to the detection of errors will be considered later.

Principle of Error Detection

The error-detecting method is based on the addition, at the originating end of the circuit, of a redundant character, known as the check character, after each checking block of text, the checking block being a typed line (about 72 characters) or portion of a line terminated by a carriage-return (CR) signal. While the principle is applicable to keyboard transmission, it will be assumed throughout this paper that transmission is from punched paper tape in a transmitter-distributor. The receiving station has a page printer which may be supplemented by a reperforator if punched tape output is wanted.

When the CR signal in the sending-station tape is sensed by the transmitter, the transmitter stops and waits while a sequence of five error-detection control characters, including the check character, are sent on the line from the error-detection circuit. The transmitter then resumes sending from the tape, the next

---

Paper 59-1106, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted June 11, 1959; made available for printing July 14, 1959.

P. H. BARRY is with Teletype Corporation, Chicago, Ill., and A. L. WHITMAN is with Bell Telephone Laboratories, Inc., New York, N. Y.

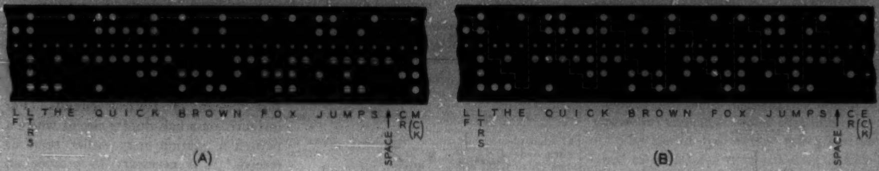


Fig. 1. Check lines

A—Horizontal
B—Spiral

character being the line-feed (LF) character necessary to advance the teletypewriter paper. The check character is a character unique to its checking block, as described in the following.

At the receiving end a new check character is locally generated from the received text by the same process as that used at the sending end. The locally generated check character is compared with the character sent over the line. If the two do not agree, it is an indication that an error has occurred in the preceding line of text. An alarm is sounded, a lamp lights, and an identifying mark is made on the copy of the line in error as described later.

The check character is generated from five odd-even counts of the marking signal elements in five separate lines along the punched tape to be transmitted. Each counting line may be a horizontal line, one of which is illustrated in Fig. 1(A). An alternate form of counting line, shown in Fig. 1(B), shifts one level vertically for each character of horizontal advance. If the tape were formed into a cylinder with level no. *5* brought adjacent to level no. *1*, this form of counting line would be a spiral; hence the term "spiral checking." Obviously, this is the helical type of spiral and not one with varying radius.

Each of the five lines generates from its count a check bit, which is a mark if there is an odd number of marks in the line, or a space if there is an even number. The check bits make five even-parity checks of the five counting lines. The check bits are then used as the five signal elements of the check character, identified as (CK) in Fig. 1(A) and (B). These figures, intended to illustrate the principle of errror detection, do not show the exact format of characters inserted at the end of each block of text. The inserted sequence is explained in detail in a later paragraph.

It will be recognized that while each parity check will detect single errors in its counting line or, in fact, any odd number of errors, it will not detect two or any even number of errors. This is due to the fact that if there is an even number of errors the odd-even count, and hence the check bit, will be the same as if there had been no errors. However, as discussed later, the chance of a double error in the chosen checking block is very much smaller than the chance of a single error so that the chance of an undetected error from double errors is correspondingly small.

The spiral checking method suggested by W. R. Young, Jr., is better than the horizontal check method in detecting errors caused by intermediate teletypewriter apparatus such as transmitters and reperforators. This is because machine errors tend to concentrate in a particular level as, for instance, when a particular transmitter sensing pin tends to stick or a particular contact gets dirty. This tends to increase the chance of double errors in a single horizontal counting line. With the spiral check a double error in one level is less likely to strike the same spiral counting line, which meets that level only once in five characters. Therefore the chance of an undetected error from a double error is reduced as much as five to one depending on the extent of concentration of errors in one level. The benefits of spiral checking may be greater where intermediate relaying is involved, as for example, on *81*-type systems and manual-tape-relay systems. However, it involves more equipment than the horizontal check.

Efficiency of Error Detection

Not enough information on the nature and occurrence of teletypewriter errors is available to permit any satisfactory estimate of the probable efficiency of a detection system. To bring out qualitatively what some of the factors are, rather than as a quantitative estimate, a theoretical calculation has been made based on some limiting assumptions.

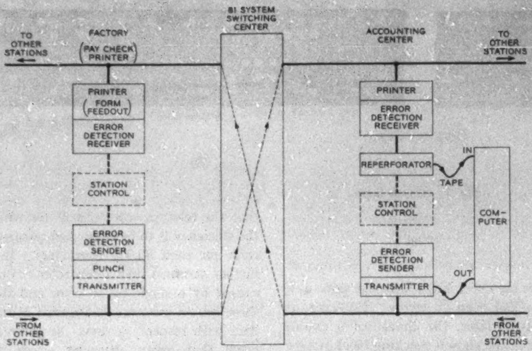
It is convenient to express error-detection efficiency in terms of the number of transmitted teletypewriter characters for one undetected error. It is assumed that the teletypewriter circuit for which the efficiency is to be computed causes an error for each 30,000 characters. It is further assumed that a character error is caused by a single bit in error and that these bits in error are distributed at random with respect to time. It is recognized that many character errors are caused by multiple-bit errors and, in fact, that some circuit disturbances, characterized as "bursts," may affect several characters; however, errors of this kind are unlikely to escape detection by the method described and are not considered here.

An undetected error, then, for the purposes of this estimate, will be the occurrence, within a single 72-character line of typing, of two errors in the same counting line. The results of the estimate are reached in two steps: 1. the probability of the occurrence of two errors in the same line of typing, and 2. the probability that they will occur in the same counting line.

To summarize the calculated results for the assumed conditions, the theoretical probability of a double error in a checking block (and hence an undetected error) is about one in 25 million characters. The five-level checking scheme used gives a further advantage of five to one, so that the theoretical probability of an undetected error is one in every 125 million characters. The formula and the method by which these figures were computed are given in the Appendix.

It is necessary to emphasize the theoretical nature of this estimate. The 125 million characters represent operation for more than a year and a half at 100 words a minute, 8 hours a day and 5 days a week. The authors do not expect such results in practice. For one thing, there is evidence that errors do not occur at random but tend to be bunched, thus increasing the probability of double errors in a single line.

It is thought significant, however, to point out that with random occurrence of errors the probability of a double, and hence undetected, error is a function of the square of the probability of a single error. Thus any improvement in transmission quality which reduces errors



TO OTHER STATIONS
FACTORY
(PAY CHECK PRINTER)
81 SYSTEM SWITCHING CENTER
ACCOUNTING CENTER
TO OTHER STATIONS
PRINTER
(FORM FEEDOUT)
ERROR DETECTION RECEIVER
STATION CONTROL
ERROR DETECTION SENDER
PUNCH
TRANSMITTER
PRINTER
ERROR DETECTION RECEIVER
REPERFORATOR
IN
TAPE
STATION CONTROL
COM-PUTER
ERROR DETECTION SENDER
OUT
TRANSMITTER
FROM OTHER STATIONS
FROM OTHER STATIONS

Fig. 2. Error detection for payroll accounting on 81 switching system

gives considerably amplified benefits in the detection of errors.

Application to an Automatic Message Switching System

The error-detection system described was planned for application to an automatic teletypewriter switching system of the *81* type.[1] Various types of data are transmitted from outlying points to a central point where there is a computer. The output information from the computer is then sent back to the originating points or distributed to other points as required. Fig. 2 shows a block diagram of a typical system. Payroll data from a factory is sent to a computer which generates paycheck information. The paychecks are printed at the factory with the information sent from the computer. Error detection is required for both directions of transmission.

There are a number of features of a message switching system which require special treatment of a superposed error-detection system. First, when a message is being transmitted there is no direct line connection between the sending and receiving stations; the message transfer at the switching center is made through two or more stages of punched tape. Thus it is not possible, during transmission, to send control signals directly from the receiving end back to the sending end. Second, no means is provided for control of the switching-center transmitter from the receiving station. Therefore, the flow of transmission is continuous and it is not possible to stop the received message in case an error is detected. This requires some means of marking the received copy so that the location of the error can be determined later For the present, it is not required that the received data go directly into the computer input; there is always a punched-tape link intervening at the receiving station.

Third, the call-directing codes (CDCs) used to direct a message through the switching system are, in general, absorbed in the switching processes and do not reach the receiver. Since the error-detection method requires comparison of the received with the transmitted material, it is necessary to exclude CDCs and confine the error-detection process solely to the text of the message. The fact that CDCs are not error-checked is not a disadvantage since the switching system itself has built-in measures to guard against losses caused by incorrect or garbled address codes.

Fourth, since the Blank character is used as a control character in the *81* switching system, the error-detection encoding system must not introduce Blanks.

Fifth, it is necessary to guard against the generation of a false disconnect signal, FIGS H, resulting from the combination of the check character, which may be any character, with adjacent control or text characters. The control-character sequence described herein accomplishes this purpose.

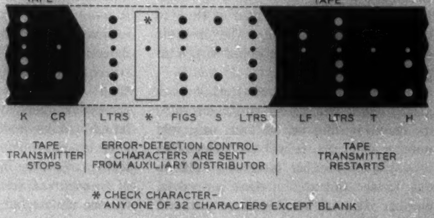
The use of this error-detection system on other kinds of switching systems or on selective calling systems requires further design of control circuits. For instance, on a half-duplex circuit where a particular station may either send or receive,[2] it is necessary to provide an automatic control of the error-detection circuit so that it will be in the send or receive condition, respectively, at the proper time.

Error-Detection Control Characters

Fig. 3 shows the format of the control characters which are inserted by the error-detection circuit at the sending end. At the end of each line of copy the originating tape will have punched in it the usual CR LF LTRS for the purpose of returning the typebox to the left edge of the paper and advancing the paper. When the CR character is read by the sensing pins of the transmitter-distributor, the sensing-pin contacts operate a circuit which opens the start-magnet circuit and so stops the transmitter. The CR character also disables the counting circuit so that the count of marks is temporarily stopped.

The five control characters are then sent over the line from an auxiliary distributor, the contacts of which are coded for the five characters in turn by a stepping selector switch. The sending of the fifth character, LTRS, closes the circuit of the start magnet and restarts the transmitter which sends the LF character in the tape followed by LTRS and the text of the next line. It also turns on the counting circuit so that the error-detection count of marks begins again with the LF character.

Table I lists in detail the things that happen at the sending and receiving ends as the various control characters are sent and received. Whenever the error-detection count of a typed line generates a Blank character as a check character, it



TAPE
TAPE
K
CR
LTRS
*
FIGS
S
LTRS
LF
LTRS
T
H
TAPE TRANSMITTER STOPS
ERROR-DETECTION CONTROL CHARACTERS ARE SENT FROM AUXILIARY DISTRIBUTOR
TAPE TRANSMITTER RESTARTS
* CHECK CHARACTER– ANY ONE OF 32 CHARACTERS EXCEPT BLANK

Fig. 3. Error-detection control characters

Table I. Functions of Control-Character Sequence*

Character	Action at Sending Station	Action at Receiving Station†
CR	Stops transmitter. Ends register count. Printer to nonprint.	Blinds reperforator. Ends register count.
LTRS	Sent through stepping selector. Timing for receiver.	Printer to nonprint condition.
Check Character	Read from register contacts and sent through stepping selector.	Check character comparison. If mismatch, printer returns to print condition, alarm lamp lights and bell rings. All registers reset to zero.
FIGS	Sent through stepping selector.	If error was detected, shifts typebox to FIGS. Prepares circuit to resume error-detection counting.
S	Sent through stepping selector.	If error, prints large gothic O over first character in the line. If no error, returns printer to print condition. Unblinds reperforator.
LTRS	Sent through stepping selector. Operates circuits to restart transmitter and to resume error-detection counting on next character. All registers reset to zero. Printer to print condition.	Timing character.

* Functions listed here are only those affecting error-detection operations. In addition, normal printing operations, such as carriage return and case shifting, take place on the appropriate characters.

† Functions are indicated opposite the characters which initiate them; in some cases, particularly those involving stuntbox contacts, operation is completed during the following character.

is converted to a LTRS character to prevent interference with the switching system. This is done at both sending and receiving ends so that the correct match occurs at the receiver. Also, the receiving end "knows" enough to discard, without action, the Blanks introduced in transmission by the transmitter start pattern of the switching system.

Receiving-End Action When Error Is Detected

With the basic premise that in the automatic switching system the flow of traffic into the receiving station cannot be stopped if an error is detected, some means are required, first, to call the attention of the operator to the fact that an error has occurred, and second, to point out where in the copy the error was located. Since the space limitations of the *28*-type printer do not favor the addition of an auxiliary mechanism to stamp a mark on the copy, it was decided to arrange the machine to print a unique teletypewriter character. Upper-case S appeared a natural choice, first, because it is not normally needed as a printing character, and second, because it is frequently used to ring a bell to attract operator attention. The typebox was equipped with a type pallet in the upper-case S position to print a large gothic O, which is large enough to encircle any other character without obliterating it.

The FIGS S signal is inserted at the sending end as a part of the error-detection control-character sequence (see Fig. 3) sent automatically after the CR signal at the end of each line. At the receiving end the printer is normally in the nonprint condition when FIGS S arrives so that no action results. In case of detected error, the printer goes into the print condition, the large gothic O is overprinted on the first character of the line in which the error was detected, the alarm lamp lights and the bell rings once.

It may seem odd that the O is printed in the same line in which the error was detected, since detection does not take place until the whole line has been printed. This results from the fact that FIGS S comes after carriage return and before line feed. Hence the typebox has returned to the left edge of the paper and prints the O before the line feed advances the paper to the next line.

The use of the regular teletypewriter bell to indicate a detected error is independent of, and does not affect, the normal bell operation at any time during transmission. Further, the normal bell signal does not light the error-detection alarm lamp.

After the attendant has noted an error detected and its location on the copy, the alarm lamp is put out by depression of the alarm-release button.

Equipment Features

The error-detection system is built exclusively to work with the *28* line of teletypewriter apparatus. The stuntbox[3] of the *28* typing unit is a basic element of the error-detection counting process and of the recognition of the series of control characters.

At the sending end, the equipment assembly is contained in the *28* automatic send-receive (ASR) set. Fig. 4 shows the laboratory prototype model. In an *81* system the keyboard is never used for sending directly on the line but only for tape punching.

The relay equipment for error detection, together with its rectifier power supply, is mounted in the base of the *28*

Fig. 4. Sending-end error-detection equipment mounted in 28 automatic send-receive set

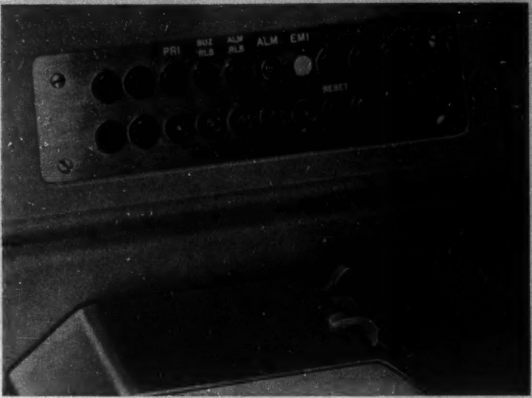


Fig. 5. Key and lamp panel at sending end

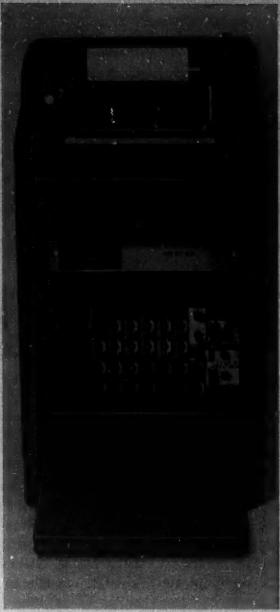


Fig. 6 (right). Receiving-end error-detection equipment in 28 receiving-only printer

Vol.
Pt.
195

ASR. Space is left here for the addition of a relay-rack panel containing a portion of the *81D1*-system station control circuit, the rest of it being in a separate apparatus cabinet.

Installation and maintenance are simplified by making all the connections to and from the panels and the machines with plug-in cords and connectors. Thus any component can be removed, if necessary, for replacement and repair.

The key and lamp control panel, mounted at a convenient place in the *28* ASR cabinet, is shown in greater detail in Fig. 5. The top row of keys and lamps is for *81*-system operation, the bottom row for error detection. The five neon lamps light up to show the operation of the flip-flop relays in the counting circuit and so change with each character during transmission. They give an indication that the error-detection circuit is operating properly and it is expected that they will be useful for maintenance. By the use of a character such as E or T, with a single marking element, the operation of the counter can be readily watched.

The receiving-end equipment is housed in the cabinet of a *28* receiving-only (RO) teletypewriter as shown in Fig. 6. The five counter lamps are shown flanking an alarm-release button which puts out the alarm lamp (yellow plastic insert on the left of the machine) when it is lighted to show a detected error. The connection of the *28* RO to the line to receive messages is controlled by the *81* station-control circuit in an auxiliary cabinet.

ERRORS WHICH THE SYSTEM WILL NOT DETECT AND WAYS TO BEAT THEM

No error-detection system of this type will detect errors made by operators in the original preparation of the tape. Reports indicate that the operator-error rate will generally be of the order of several times the transmission-error rate. A possible way to check for operator errors with this system is to compare the copy printed while the message is being transmitted with the original source of the copy from which the tape was punched.

This sort of comparison check will also detect another possible source of errors, those due to incorrect operation of the transmitter-distributor. Since the error-detection system is actuated by line signals, it gives assurance only that the signals sent on the line are correctly reproduced at the receiving end. If the sending transmitter has either sensing-pin or contact failure, this will not be detected. However, a copy check as suggested will catch errors from this cause.

In the process of planning the design, means of automatically detecting errors caused by the transmitter were considered. One plan required a double reading of the sending-end tape, but such arrangements were felt to be too expensive and too cumbersome for operator use. If experience indicates that detecting machine errors is vitally important, further design work will be needed.

The spiral checking method gives a high degree of protection against errors caused by intermediate machines such as switching-system reperforator-transmitters, or, for manual relay, reperforators and the associated transmitters.

Another source of errors which will not be detected by the system are those caused by machine failure of the reperforator at the receiving end. Again what is checked are the line signals reaching the reperforator. If the punch mechanism, for instance, does not punch the holes called for by the line signals, this is not detected. It is anticipated that with proper maintenance, errors of this kind will be so infrequent as not to be troublesome.

Conclusions

The reduction in undetected errors by the method described in this paper is expected to lead to significant improvement in the quality of teletypewriter transmission. This is particularly needed where the type of material transmitted, as, for example, numerical digits, is such that errors cannot be readily seen by visual inspection.

Appendix. Theoretical Probability of Double and Hence Compensating Bit Errors in a Line of Typing

The probability, P, that an event will happen exactly r times in n trials is given by the formula:[4]

$$P=\frac{n(n-1)\ldots(n-r+1)p^r q^{n-r}}{r!}$$

where

P = probability of r errors in n characters
n = number of characters in a typed line, here taken as 72
r = number of errors in a line of n characters
p = probability of a single error, here taken as one error in 30,000 characters
q = probability that an error does not occur; equal to $(1-p)$

For this particular illustrative example it is assumed that a character error is caused by a single bit in error and that these bits in error are distributed at random with respect to time. An undetected error, then, for the purpose of this estimate, will be the occurrence, with a single 72-character line of typing, of two errors in the same counting line.

When r is 2, the equation reduces to:

$$P=\frac{n(n-1)p^2q^{n-2}}{2}$$

Since p is very small, q is very nearly 1, hence q^{n-2} is nearly 1. Then:

$$P=\frac{(n^2-n)p^2}{2}$$

For

$n=72$ and $p=1/30{,}000$
$P=1/352{,}000$

This means that the probability of occurrence of a double error is once in every 352,000 typed lines, or, with 72 characters in a line, once in about 25,000,000 characters.

The method of error detection here described gives an additional five-to-one advantage because the chances are only one in five that the two errors in a typed line will fall in the same counting line. So, finally, the theoretical probability of an undetected error for the assumed conditions, is one in 125,000,000 characters.

A similar calculation could be made for $r=4$. However, the probability of an undetected error from four errors in one checking line is enormously smaller than the probability from two so that it can be neglected in comparison.

References

1. A Full Automatic Private-Line Teletypewriter Switching System, W. M. Bacon, G. A. Locke. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 473–80.

2. The 83B1 Teletypewriter Selective-Calling System, C. W. Smith, C. J. Votaw, A. L. Whitman. *Ibid.*, pt. I (*Communication and Electronics*), vol. 78, May 1959, pp. 119–24.

3. Simplified Printing Telegraph Switching and Integrated Data Processing, J. B. Booth, R. H. Klich. *Ibid.*, vol. 75, Mar. 1956, pp. 11–14.

4. Introduction to Probability Theory and Its Applications (book), William Feller. John Wiley & Sons, Inc., New York, N. Y., second edition, 1957, chap. 6, p. 137.

5. Error-Checking for 5-Channel Telegraphic Tape, R. A. Barbeau. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, May 1958, pp. 190–93.

6. Optimum Block Length for Data Transmission with Error Checking, F. B. Wood. *Ibid.* (Jan. 1959 section), pp. 855–61.

7. Some Aspects of Telegraphic Data Preparation and Transmission, William B. Blanton. *Western Union Technical Review*, New York, N. Y., Oct. 1957, pp. 142–51.

Considerations in SSB and ISB Systems for Long-Distance Radiotelegraph Communications

WALTER LYONS
NONMEMBER AIEE

SINGLE side-band (SSB) and independent side-band (ISB) radiotelegraph systems allow for a high degree of frequency division multiplexing as contrasted to the requirement for time division multiplexing in usual frequency shift keyed carrier systems. This advantage is the principal justification for its use, especially for those circuits requiring high-speed traffic-handling capacities.

For the transmission of bit elements smaller than 10 milliseconds, the multipath delay distortion may be high to a degree that is intolerable for a commercial circuit. To reduce this distortion, it is required that the keyed pulse be "squared up," in other words, that it attain keying harmonics. This, in effect, forces the employment of large bandwidths of emission and reception. Thus, frequency conservation is defeated and power is wasted.

For smaller bit lengths, say, under 5 milliseconds, it may be impossible to improve reception by the method of adding harmonics to the keying pulse. For this reason, it has been proposed to change the code arrangement from sequential bit information to simultaneous frequency-separated bits. Simultaneous code presentation has the additional advantage in error-correcting equipment, where changes in code are usual, to reduce the amount and cost of equipment.

In the application of SSB for multichannel transmission, it is important, because of limited power which is available at this time from linear amplifiers, to limit the peaks of modulation. If the individual tone channels are keyed in an on-off manner, phase locking these tones is sufficient to assure a small peak-to-average amplitude ratio. For frequency-shifted tone channels, other means such as compatible SSB may be employed instead. On the other hand, and for greater simplicity, increasing the number of these channels will reduce the time duration of high amplitude peaks and render them negligible.

Because tone channeling on SSB transmission can be very accurately controlled and spaced because of the low audiofrequencies used, frequency shifting at optimum values can be employed to obtain the highest effective power in the transmission bandwidth. In this way, over 40 teleprinter channels can be accommodated in a voice band at transmitter power levels well within present linear amplifier techniques.

Recent developments in transistorization and semiconductor devices greatly reduce heat dissipation, size, and weight of a great part of the SSB receiver, which would be hardly possible under the older types. Due to these newer tools, maintenance is greatly reduced. The net result is the ability to expand

---

Paper 59-1091, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted November 25, 1958; made available for printing June 25, 1959.

Walter Lyons is with RCA Communications, Inc., New York, N. Y.

The author wishes to express his appreciation to J. C. Hepburn, for providing information regarding propagation requirements, to A. B. Bodonyi, for much help in compilation of data and construction of the curves of Figs. 1 and 2; to D. S. Rau, E. D. Becken, and Dr. H. O. Peterson, for their editing and suggestions; and to Tele-Signal Corporation for their co-operation in furnishing Figs. 3 and 4.

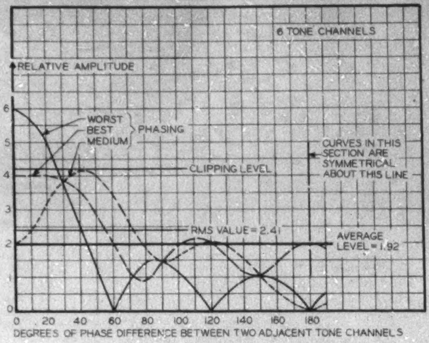


Fig. 1. Amplitude characteristic for six tones

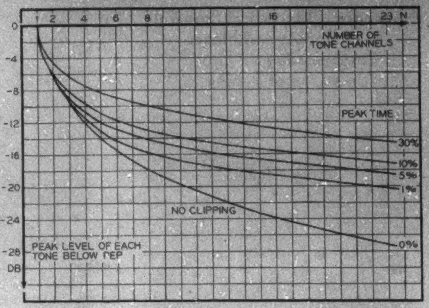


Fig. 2. Multitone levels for various clipping periods

operational circuits without the usual requirements for increase in personnel, building space, and real estate.

Background

The application of SSB techniques to point-to-point long-distance radiotelegraph communication circuits promises the greatest growth possibilities in this field since its inception. However, in order to realize the benefits to be derived, every advantage offered by SSB must be thoroughly exploited, and all devices which can reduce its shortcomings should be employed.

The advantages accruing to SSB working for point-to-point circuits at high frequency, as compared with ordinary carrier frequency shift (FSK), reside in economy in equipment and antennas, greater simplicity in operation, spectrum conservation, and power gain. The practical realization of economies in equipment depends upon adequate telegraph circuit traffic loading to justify the initial investment and change from the older FSK gear. Transmitted frequency spectrum compression and power gain go hand in hand and are the peculiar attributes of SSB, due to its inherent frequency stability because of carrier reference at audiofrequencies. This is because the frequency-shifted channels being referred to the carrier are low audiofrequencies, and problems in maintaining small shifts and accurate centers are far easier than at carrier shift frequencies.

System Specification

Before further considerations, it is appropriate to set forth the basic requirements for a commercial point-to-point long-distance radiotelegraph circuit. They are that the circuit operate at least 95% of the time for 24 hours daily throughout the sunspot cycle, and that the error rate during operation of the radio path be less than 0.01%. To assure this performance, it has been determined empirically that about a 5-kw input to a 13-decibel rhombic should be provided for each kc emission bandwidth, assuming two and occasionally three frequency changes per day. This does not apply to circuits whose great circle paths go through the high absorption auroral zones.

Implementation

Present practice in loading a 3-kc side band of an SSB transmitter with three time-division 4-channel multiplex aggregates thus will require 15-kw average power and 90-kw instantaneous peak envelope power. For a fixed bandwidth-per-tone channel the average radiated power required increases directly as the number of subcarriers, the peak power rises as the square. Clearly, means for limiting peak power is indicated in order that economic considerations do not lead to a decision against SSB working for this degree of traffic loading. Conventional means for limiting the peak radio-frequency power, however, will result in an increase in spurious signal level with attendant signal degradation. There are two ways to remedy this situation: one would be to limit the modulation peaks, which requires, for minimum spurious output, a compatible SSB generator because the radio-frequency envelope is a replica of the modulation input; the other way is to increase the number of input tone channels so that the duration of each peak is reduced to about 10% of a signal element. Assuming 10-millisecond signal elements, 1-millisecond peak duration is tolerable and indicates a minimum requirement for approximately six tone channels. The assumption is that all channels are not resting on fixed tones and channeling tones are 120 cps (cycles per second) or more apart. This is a fortunate circumstance, since multipath delay distortion effects are not excessive when a bit of at least 10 milliseconds is used. If the three 4-channel multiplex aggregates are split up into six diplex (2-channel) channels (12-teleprinter capacity), peak power will quadruple, but chances are that the duration of any one peak will be short enough nearly to be negligible. The curves for various phasing conditions in Fig. 1 show that, for 120-cps separation, clipping or limiting by the transmitter above the nominal peak level 4, the duration is 1.23 milliseconds. The value decreases as the tone separation increases. Fig. 2 presents the arguments for peak clipping of a large number of tone channels. The equivalent power gain due to clipping increases markedly with the number of tone channels. The percentage time of clipping is a measure of the degree of spurious radiation or distortion. For the 5% and smaller curves, no single peak duration exceeds 1 millisecond for most unfavorable tone phasing. For many subchannels, each being frequency shift keyed, the instantaneous envelope amplitude approaches the probability density of a Rayleigh distribution. The expected distortion therefore will be considerably less than the curves indicate.[1]

A reduction in emission bandwidth and consequent increase in effective power is now feasible, since at the reduced keying speed the transmitted pulse

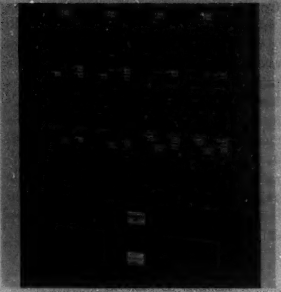


Fig. 3. Section of transmitter FSK subchanneling. Standard 84-inch rack houses 48 channels

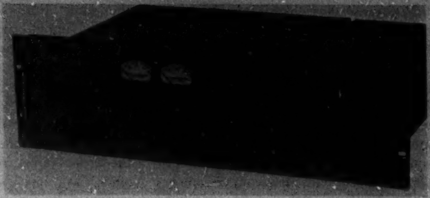


Fig. 4. FSK subchannel diversity receiver and tone transmitter

need not be trapezoidal for protection against multipath delay distortion. The transmission of a sine-wave-shaped signal element will serve to reduce interchannel crosstalk, allowing the use of filters having smooth cutoff characteristics with inherently low phase distortion. In this manner, a system may be devised to provide as many as 23 subcarrier frequency-shifted channels, each operated at diplex speed, accommodating 46 60-words-per-minute teleprinters in a single SSB voice channel. Required parameters are a modulation index of 0.8 and subcarrier channel separation 120 cps, starting at 300 cps and extending to 2,940 cps. A short-term traffic test of a portion of such a system, as compared with the usual 4-channel multiplex (quad) channeling at 340-cps shift and 680-cps channel separation, has shown marked improvement in error rate. For the comparison, four diplexed channels and two quad channels were operated on an SSB voice channel. If, now, a circuit is considered with requirements for only 12 teleprinters, the total bandwidth from carrier need only be about 900 cps and side-band power output roughly a quarter of the power required for full occupancy of the voice band. This is a power level which may be served adequately by linear amplifiers. Heavily loaded circuits and those carrying services such as phone or facsimile in addition might consider using envelope elimination and restoration SSB transmitters not to be limited in power for reasonable levels of spurious radiation.[2]

Frequency-shift keying has been plagued by multipath delay distortion to such an extent that FSK exciters have been built with switches to change transmission to on-off keying. It was possible to overcome the elongation of the "mark" signal by reducing keying weight on transmission. It is an unfortunate circumstance that misconception, due probably to frequency-modulation broadcasting techniques, put the point-to-point radio communications systems on a wide-shift basis. Perhaps also, in an attempt to overcome selective fading effects, the superimposed 200-cps phase modulation dictated the 850-cps shift even for solo channels.

Considerable evidence on long-haul circuits supports the thesis that a reduction in shift reduces jitter. It may be that selective fading is reduced to such an extent that multipath delay tends to act equally on both mark and "space" frequencies. This should facilitate the reduction of distortion in the final signal by simple adjustment of the phase of inspection of the signal element, approaching in this manner an unbiased (equally biased) signal. Thus the reduction in shift conforms to the characteristics of the keying mode, polar and unbiased, with attendant advantages in error reduction.

It is recognized that, in receiving signals with considerable multipath delay, wider filters will reduce the resultant jitter, but the reduction in frequency shift below 100 cps, serving to equalize the delay of both mark and space signals, allows for inspection closer to the trailing edge of the first received element, at the center of the unmutilated resulting bit. Including the increase in power attending the reduction in keying speed permitting sine-wave keying, the case for an FSK modulation index below unity is summarized.[3]

SSB working can enable simplification in terminal equipment by the facility for trading frequency division for time division multiplexing. Most error-correcting systems used in point-to-point radio communications at present employ the 7-unit Moore Code for error detection. This code invariably denotes a character by 3-mark and 4-space elements. Any deviation from this ratio is an error, calling for repetition by the sender. Most land line traffic uses the 5-unit start-stop code, requiring, therefore, a code conversion device for operating into radio circuits protected by error-correcting systems. For this purpose it is required that information for each conversion be changed from serial code to simultaneous presentation of the code elements. By the use of separate tones each to denote an element of the sequential code, each such tone being keyed simultaneously for mark or space, simultaneous code presentation is effected very simply, resulting in equipment reduction and signal element elongation, if desired, without sacrificing traffic handling speed.[4]

Data handling other than regular message should become, very soon, a major service on long-haul circuits. Accommodation of codes peculiar to a specific data-handling system can mean little more complication than adding additional tone channels for simultaneous keying. There is one pitfall, however, that must be avoided and which may be overlooked easily. The time delay in most filters and transfer lines varies with frequency. Steps must be taken, therefore, to assure equal delay for all the filters in a simultaneous code group, as well as the transfer line, over the frequency range of the group. This is an additional item of expense which must be included in an economic analysis of the two code types. Moreover, the need for equipment reliability in each of these tone channels is far greater than in the equipment for sequential keying.

The versatility in point-to-point services of independent side-band working is a great economic argument for this mode. Assuming sufficient power, phone, facsimile, and many telegraph channels may be operated at the same time. These services may be handled at minimum bandwidth occupancy, and should simplify frequency allocation by block assignment. For this type of operation, spurious signal radiation should be kept as low as is practicable, not alone to reduce interference out of band, but also self-interference in band, to which facsimile is especially vulnerable. For this purpose spurious signals should be down 30 decibels or more as measured by the standard 2-tone test.

Multiservice operation naturally is accompanied by equipment complication. Especially is this so when diversity reception techniques are used. The situation breeds, in the designer, a desire to simplify, sometimes at the sacrifice of performance. This tendency is understandable, especially in view of pressure from those who have the responsibility for maintaining the equipment. One design example reduced the problem by divorcing the receiver from the so-called terminal gear (low-frequency equipment), installing the audiofrequency components in the central office. For large commercial carriers, this has two shortcomings: first, the expense of doubling control line facilities for diversity reception, and second, the placing of an electronic maintenance load on an organization having little electronic experience. Diversity at the central office requires that the transfer lines have a very large dynamic range; otherwise, fading signals drop into the noise level suffering serious degradation.

Fortunately, recent developments in transistorization and semiconductor devices allow the possibility of very greatly reducing heat dissipation, the size and weight of receiver units, terminal equipment, and some transmitting units; at the same time effecting considerable circuit simplification. The net result is the ability to expand operational circuits without the usual requirement for expansion in personnel, building space, and real estate. Under such circumstances, it is questionable whether it is economical to consider automation in the form of direct control from the central office. It appears that a minimum personnel requirement at radio stations for maintenance and signal analysis can perform operating tasks as required without intruding on primary responsibilities. Extreme caution is indicated before application of servos and other schemes for automation lest the design result in a requirement for greater human effort in maintenance and service.

Economy in space due transistorization of transmitter channeling equipment is illustrated in Fig. 3. This is an assembly of 32 frequency shift tone keyers (transmitters) and 32 tone signal converters (signal demodulators), making a complete terminal facility for 32 FSK subchannels on SSB-ISB transmission.

Fig. 4 is a complete assembly for tone channel diversity reception, including a tone keyer for transmitting the received signal to the central office. The four separate units comprise two subchannel receivers and the diversity comparator between them, with the tone keyer on the extreme right.

Conclusions

With the upcoming low in the sunspot cycle, frequency conservation will be a prime necessity. The availability of high-power SSB-ISB transmitters, together with techniques outlined above, enables the industry to increase traffic at the same time that frequency occupancy is reduced. Justification for changing over to SSB-ISB might be the establishment of overseas trunk circuits or forked operation to points close to the same azimuth.

The additional complications introduced require greater co-operation between correspondents than heretofore in operating simple carrier FSK. The establishment of standards should help, and efforts in this direction should be made soon to cope with what is expected in propagation in the near future.

References

1. Compatible Single Side Band in HF Communications, **Kenneth B. Boothe.** *Aviation Age*, New York, N. Y., Sept.-Oct. 1957.

2. Comparison of Linear SSB Transmitters with Envelope Elimination and Restoration SSB Transmitters, **Leonard R. Kahn.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 44, no. 12, Dec. 1956.

3. Design Considerations for Frequency Shift Keyed Circuits, **Walter Lyons.** *RCA Review*, Princeton, N. J., June 1954.

4. Printing Telegraph System, **John B. Moore, Richard E. Mathes.** U. S. Patent no. *2,183,147*, Dec. 12, 1939.

Criteria and Tests for Realizability of the Inductance Matrix

YILMAZ TOKAD
NONMEMBER AIEE

MYRIL B. REED
NONMEMBER AIEE

ENERGY considerations and experience have both indicated that measurements of the *R*, *L*, and *C* coefficients of an existing electrical network will lead to the positive semidefinite quadratic forms[1-3]

$$\mathfrak{I}^T(t)\mathfrak{R}\mathfrak{I}(t)\geq 0 \qquad (1)$$

$$\mathfrak{Q}^T(t)\mathfrak{D}\mathfrak{Q}(t)\geq 0 \qquad (2)$$

$$\mathfrak{I}^T(t)\mathfrak{L}\mathfrak{I}(t)\geq 0 \qquad (3)$$

Given a particular numerical set of symmetrical matrices which are to be used as the coefficient matrices $\mathfrak{R}$, $\mathfrak{D}$, $\mathfrak{L}$ of the preceding three inequalities, the positive semidefiniteness of the coefficient matrices is only certain for the first two. This certainty is a consequence of the matrices $\mathfrak{R}$ and $\mathfrak{D}$ being diagonal with nonnegative entries on the diagonal. The inductance (self-and mutual) matrix, $\mathfrak{L}$ of equation 3, on the contrary, may have non-zero off-diagonal entries with positive and negative signs in a variety of patterns. Cauer gave a realization technique for an $\mathfrak{L}$ matrix, satisfying equation 3, by using perfect coupled ($k=1$) transformers.[3] Therefore a particular numerical matrix (symmetric and real) can be realized by a set of coupled coils if it is positive semidefinite. The positive definiteness or semidefiniteness of such a matrix can be tested by checking the nonnegativeness of all the principal subdeterminants of the given matrix. The arithemtic involved in this multiple evaluation of determinants can readily become extensive.

In this paper a different test, in terms of the eigenvalues of the inductance matrix, is formulated. Indications are that this eignevalue test method involves appreciably less arithmetic than the multiple determinant method. Certain necessary and sufficient conditions on the eigenvalues as well as on the coefficient-of-couplings are presented as of interest in their own right as well as incidentals to the eigenvalue test metod.

That $|k_p|\leq 1$ Is Necessary for the Positive Semidefiniteness of the Inductance Matrix

Consider the proposition: If $|k_{pq}|\leq 1$, then $\mathfrak{L}$ is positive semidefinite. This

---

Paper 59-1090, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted January 29, 1959; made available for printing July 20, 1959.

Yilmaz Tokad and Myril B. Reed are with Michigan State University, East Lansing, Mich.

theorem is actually not true, i.e., not sufficient. The necessary condition for this theorem is proved by showing: If $\mathcal{L}$ is positive semidefinite, then $|k_{pq}|\leq 1$. This necessary condition is proved next.

Let the inductance matrix be reshaped to exhibit the coefficients of coupling. Thus

$$\begin{bmatrix} \sqrt{L_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{L_{22}} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sqrt{L_{nn}} \end{bmatrix} \begin{bmatrix} 1 & k_{12} & \dots & k_{1n} \\ k_{21} & 1 & \dots & k_{2n} \\ \vdots & \vdots & & \vdots \\ k_{n1} & k_{n2} & \dots & 1 \end{bmatrix} \times$$

$$\begin{bmatrix} \sqrt{L_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{L_{22}} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sqrt{L_{nn}} \end{bmatrix} = \mathcal{L}_d \mathcal{K} \mathcal{L}_d \quad (4)$$

where the coefficients of coupling, $k_{pq}=k_{qp}$, are

$$k_{pq}=\frac{L_{pq}}{\sqrt{L_{pp}L_{qq}}} \quad (5)$$

Since the diagonal matrices, $\mathcal{L}_d$, are nonsingular, the positive semidefiniteness of $\mathcal{K}$ implies the positive semidefiniteness of $\mathcal{L}$ and vice versa.

Assume now that $\mathcal{L}$ is positive semidefinite, i.e., that equation 3 is valid for all $\mathcal{I}(t)$. In particular, then, equation 3 is valid if all entries in $\mathcal{I}(t)$ are zero but any two—$i_p(t)$ and $i_q(t)$. If $p\neq q$, equation 3 is in detail, since $L_{pq}=L_{qp}$,

$$L_{pp}i_p^2(t)+L_{qq}i_q^2(t)+2L_{pq}i_p(t)i_q(t)\geq 0 \quad (6)$$

But the positive semidefiniteness of this quadratic form compels

$$L_{pq}^2\leq L_{pp}L_{qq} \quad (7)$$

Hence, if equation 3 is valid so is equation 7 which requires by equation 5

$$|k_{pq}|=\frac{|L_{pq}|}{\sqrt{L_{pp}L_{qq}}}\leq 1 \quad (8)$$

Therefore, since equation 3 implies equation 8, the relation 8 is necessary.

That $|k_{pq}|\leq 1$ Is Not Sufficient for Positive Semidefiniteness of the Inductance Matrix

That $|k_{pq}|\leq 1$ is not a sufficient condition for positive semidefiniteness of $\mathcal{L}$ is demonstrated next, by showing that equation 8 is not sufficient for equation 3 for the third-order inductance matrix although it is for the second-order case.

First, for a second-order matrix, $\mathcal{K}_2$, where

$$\mathcal{K}_2=\begin{bmatrix} 1 & k_{12} \\ k_{21} & 1 \end{bmatrix} \quad (9)$$

$\det \mathcal{K}_2=1-k_{12}k_{21}\geq 0$ for $|k_{12}|=|k_{21}|\leq 1$
and $\mathcal{K}_2$ is a positive semidefinite matrix and vice versa. Hence for this second-order case the usual condition of the coefficients of couplings, $|k_{pq}|\leq 1$, is necessary and sufficient for positive semidefiniteness of $\mathcal{K}_2$ and so of $\mathcal{L}$.

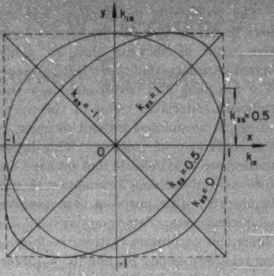


Fig. 1. Loci of k_{pq} for three windings

But consider the third-order matrix,

$$\mathcal{K}_3=\begin{bmatrix} 1 & k_{12} & k_{13} \\ k_{21} & 1 & k_{23} \\ k_{31} & k_{32} & 1 \end{bmatrix} \quad (10)$$

This matrix may or may not be a positive semidefinite matrix with the restrictions, $|k_{pq}|=|k_{qp}|\leq 1$. Indeed, if $k_{12}=k_{21}=0.5$, $k_{13}=k_{31}=-0.5$, and $k_{23}=k_{32}=0.6$, $\det \mathcal{K}_3<0$ and $\mathcal{K}_3$ is neither positive definite nor positive semidefinite. Therefore, the requirement $|k_{pq}|\leq 1$ is not sufficient for the positive definiteness or semidefiniteness of $\mathcal{K}_3$.

The complete pattern of the restrictions on the k_{pq} for the third-order case can be formulated as follows: let $x=k_{12}=k_{21}$, $y=k_{13}=k_{31}$, $z=k_{23}=k_{32}$. The determinant of $\mathcal{K}_3$ is, then,

$$\det \mathcal{K}_3=1+2\times yz-x^2-y^2-z^2 \quad (11)$$

It can be shown that $\mathcal{K}_3$ is positive definite for all points in a volume enclosed by the surface

$$1+2\times yz=x^2+y^2+{}^2z \quad (12)$$

and is positive semidefinite for all points on this surface.

Fig. 1 represents the projection on the $x-y$ plane of the intersection curve of this surface by a horizontal plane $z=k_{23}$. Fixing, say, k_{qr}, restricts the range of values of the other coefficient of coupling, k_{pq} and k_{pr}, for a positive semidefinite $\mathcal{K}_3$. For example, let $p_i(k_{pq}, k_{pr})$ be a point on the $x-y$ plane. If p_i lies inside the ellipse with k_{qr}=constant, then $\mathcal{K}_3$ is positive definite. If p_i is on the ellipse, then $\mathcal{K}_3$ is positive semidefinite, and if p_i is outside the ellipse but inside square of sides 2, then $\mathcal{K}_3$ is neither positive definite nor semidefinite.

If two of the coupling coefficients, k_{pq}, k_{pr}, are fixed, then the third one, k_{qr}, cannot be chosen arbitrarily to maintain the positive semidefiniteness of the matrix $\mathcal{K}_3$. Indeed, in this case it can be seen that k_{qr} must satisfy the relation

$$(k_{qr})_1\leq k_{qr}\leq (k_{qr})_2$$

where

$$(k_{qr})_1, (k_{qr})_2=k_{pq}k_{pr}\pm \sqrt{(1-k_{pq}^2)(1-k_{pr}^2)} \quad (13)$$

Therefore, for the positive definiteness or semidefiniteness of $\mathcal{K}_3$ the following conditions are necessary and sufficient:

$$|k_{pq}|\leq 1 \quad (14A)$$

$$k_{12}^2+k_{13}^2+k_{23}^2-2k_{12}k_{13}k_{23}\leq 1 \quad (14B)$$

In this case, since condition 14(B) contains the condition 14(A), then the necessary and sufficient condition is given by equation 14(B).

The preceding third-order example shows that the positive definiteness or semidefiniteness of the coefficient-of-coupling matrix, and so the inductance matrix, is more complex than simply specifying that all coefficients-of-coupling be less than unity in magnitude. The complexity of the restrictions increases rapidly as the order of the matrix $\mathcal{K}$ increases.

Necessary and Sufficient Requirements on the Inductance Matrix in Terms of Its Eigenvalues

Consider now the eigenvalue formulation. Necessary and sufficient conditions that any inductance matrix of any order is positive definite or semidefinite can be formulated in terms of the characteristic roots, or eigenvalues, of this matrix.[4] These eigenvalues are the roots, λ_i, of the characteristics polynomial of a matrix $\mathcal{L}$, defined by

$$P(\lambda)=\det[\mathcal{L}-\lambda\mathcal{U}]=(-1)^n[\lambda^n-\alpha_1\lambda^{n-1}+\alpha_2\lambda^{n-2}-\dots(-1)^n\alpha_n]$$
$$=(-1)^n(\lambda-\lambda_1)(\lambda-\lambda_2)\dots(\lambda-\lambda_n) \quad (15)$$

where $\mathcal{U}$ is the identity, or unit, matrix of order n and where the coefficients α_i of this characteristic polynomial are

(*a*) α_i=sum of all possible different products of the eigenvalues taken i at a time.

(*b*) α_i=sum of the diagonal minors of order i of $\mathcal{L}$.[5] Particularly $\alpha_n=\det \mathcal{L}$.

Applying Descartes rule of signs[6] to equation 15 leads to

Lemma: If and only if $\alpha_i\geq 0$, then $\lambda_i\geq 0$

and then to

Theorem: If and only if $\mathcal{L}$ is real, symmetric, and positive definite (semidefinite), then the eigenvalues, λ_i are real and $\lambda_i>0$ $(\lambda_i\geq 0)$[7,8]

Also, from the lemma, α_i are real and $\alpha_i > 0$ ($\alpha_i \geq 0$)

This theorem indicates an inspection technique on the coefficients, α_i, or the roots, λ_i, of the characteristic polynomial to detect positive definiteness or lack of it. However, in even the recent past this test would be essentially useless because of the formidable arithmetic barrier to calculating either the α_i or the λ_i. The digital computer has altered this situation to such an extent that it is a matter of minutes or even seconds, to now calculate the eigenvalues of a matrix directly to check for positive definiteness or semidefiniteness; or the multideterminant method is indicated by item *b*.

For example, the matrix

$$\mathcal{L} = 10^{-6}\begin{bmatrix} 0.2240 & 0.2691 & 0.2691 & 0.1724 & 0.1039 & 0.1039 \\ 0.2691 & 0.4713 & 0.3851 & 0.3163 & 0.2333 & 0.08629 \\ 0.2691 & 0.3851 & 0.4713 & 0.3163 & 0.2684 & 0.2333 \\ 0.1724 & 0.3163 & 0.3163 & 0.6188 & 0.4092 & 0.4092 \\ 0.1039 & 0.2333 & 0.2684 & 0.4092 & 1.048 & 0.7814 \\ 0.1039 & 0.08629 & 0.2333 & 0.4092 & 0.7814 & 1.048 \end{bmatrix}$$

recently was calculated from certain approximation formulas, for the calculation of the air-gap inductances, at a particular instant, of all the commutating coils in a d-c machine, and the question arose as to whether this matrix is an inductance matrix, i.e., does it correspond to a possible set of coupled coils, or mathematically is it positive semidifinite?[9] In a matter of about one minute a digital computer (MISTIC) gave the eigenvalues

$\lambda_1 = 0.0312 \times 10^{-6}$
$\lambda_2 = 0.0666 \times 10^{-6}$
$\lambda_3 = 0.2429 \times 10^{-6}$
$\lambda_4 = 0.2960 \times 10^{-6}$
$\lambda_5 = 0.9102 \times 10^{-6}$
$\lambda_6 = 2.3346 \times 10^{-6}$

through use of one of the programs in the computer library. All of these eigenvalues are real and positive, hence the matrix, $\mathcal{L}$, is positive definite and so an allowable inductance matrix.

The coefficients of coupling, k_{pq}, are used in the foregoing largely in the negative sense of showing that an easily stated necessary condition is not sufficient. The foregoing and useful necessary and sufficient criteria in terms of the eigenvalues, is formulated without recourse to the coefficients-of-coupling. This latter pattern seems to indicate that perhaps the coefficient-of-coupling formulation should be used for other purposes than positive semidefiniteness determinations. To clinch this conclusion further the following relations among α'_i and the k_{pq} have been established. Where α'_i are the coefficients of the characteristic equation of the matrix, $\mathcal{K}$, i.e.,

$$\det[\mathcal{K} - \lambda\mathcal{U}] = (-1)^n[\lambda^n - \alpha'_1\lambda^{n-1} + \alpha'_2\lambda^{n-2} - \ldots (-1)^n\alpha'_n] = 0$$

where again α'_i have the same properties as α_i. From characterization (b) following equation 15, the α'_i are

$$\alpha'_1 = \sum_{p=1}^{n} k_{pp} = n > 0 \quad (\text{trace of } \mathcal{K})$$

which is always positive,

$$\alpha'_2 = \sum_{p<q}^{n} (k_{pp}k_{qq} - k_{pq}^2) = \sum_{p<q}^{n} (1 - k_{pq}^2) > 0$$

i.e., the point $P(k_{12}, \ldots, k_{1n}, k_{21}, \ldots, k_{(n-1\ n)})$ must be in the $n(n-1)/2$ dimensional sphere of radius $\sqrt{n(n-1)/2}$

$$\alpha'_3 = \sum_{p<q<r}^{n} [1 + 2(k_{pq}k_{pr}k_{qr}) - (k_{pq}^2 + k_{pr}^2 + k_{qr}^2)] > 0$$

$$\begin{aligned}\alpha'_4 = \sum_{p<q<r<s}^{n} [1 - (&k_{pq}^2 + k_{pr}^2 + k_{ps}^2 + k_{qr}^2 + k_{qs}^2 + k_{rs}^2) \\ &+ 2(k_{pq}k_{pr}k_{qr} + k_{pr}k_{ps}k_{rs} + k_{pq}k_{ps}k_{qs} + k_{qr}k_{qs}k_{rs}) \\ &+ (k_{pq}^2k_{rs}^2 + k_{pr}^2k_{qs}^2 + k_{ps}^2k_{qr}^2) \\ &- 2(k_{pq}k_{pr}k_{qs}k_{rs} + k_{pq}k_{ps}k_{qr}k_{rs} + k_{pr}k_{ps}k_{qr}k_{qs})] > 0\end{aligned}$$

$$\vdots$$

$$\alpha'_n = \det \mathcal{K} > 0$$

The complete detailing of this general set of inequalities is not feasible although these inequalities could be established, of course, for a particular case, or for many cases, the k_{pq} evaluated from the *L*-values, and the whole tested for positiveness. Or it should be possible to set a computer to work locating and evaluating the appropriate minors of $\mathcal{K}$ and then combining them. In contrast with the clearly simple computer eigenvalue pattern, there seems little chance that the coefficient-of-coupling technique can compete.

References

1. NETWORK ANALYSIS AND FEEDBACK AMPLIFIER DESIGN (book), **H. W. Bode.** D. Van Nostrand Company, Inc., Princeton, N. J., 1956.
2. NETWORK SYNTHESIS (book), **D. F. Tuttle.** John Wiley & Sons, Inc., New York, N. Y., 1957.
3. SYNTHESIS OF LINEAR COMMUNICATION NETWORKS (book), **W. Cauer.** McGraw-Hill Book Company, Inc., New York, N. Y., 1958.
4. A SURVEY OF MODERN ALGEBRA (book), **G. Birkhoff, S. MacLane.** The Macmillan Company, New York, N. Y., 1957.
5. AN INTRODUCTION TO THE THEORY OF CANONICAL MATRICES (book), **H. W. Turnbull, A. C. Aitken.** Blackie and Son, Ltd., London, England, 1932.
6. NUMERICAL ANALYSIS (book), **Kaiser S. Kunz.** McGraw-Hill Book Company, Inc., 1957.
7. LINEAR ALGEBRA FOR UNDERGRADUATES (book), **D. C. Murdoch.** John Wiley & Sons, Inc., 1957.
8. ELEMENTARY MATRIX ALGEBRA (book), **Franz E. Hohn.** The Macmillan Company, 1958.
9. SWITCHING NETWORKS FOR COMMUTATING MACHINES, **H. E. Koenig H. K. Kesavan.** *Report No. 2*, Engineering Experiment Station, Michigan State University, East Lansing, Mich.

Design Features of a New Frequency Shift Carrier Telegraph System

M. L. STEPHENS
ASSOCIATE MEMBER AIEE

NARROW-BAND frequency shift carrier telegraphy has found steadily increasing usage throughout the communications industry since such systems were first introduced. Today, the major communication companies of the United States and Canada are rapidly expanding their carrier telegraph networks, and many applications are being found among such industries as the railroad and pipeline companies. A surprising number of data transmission problems may be solved through the use of these modern high-speed telegraph circuits.

This paper describes some of the design features of a new carrier telegraph system designated as type *23A*. The broad design objective was to develop a telegraph system with the rigid performance requirements and flexible application demanded by the user, making use of the latest developments in components and manufacturing technology.

A complete *23A* terminal of 18 channels occupies about 7 feet of space on a standard 19-inch rack; the over-all depth is 10 inches. The circuits are transistorized, and the full terminal dissipates about 55 watts of power exclusive of loop drains. Comparable vacuum-tube equipment dissipates nearly ten times this power. The transistor circuits operate on negative 48-volt office battery, and loops may be operated on the same office battery or on the usual telegraph supplies. Loop facilities are built in, and the channels operate with most existing neutral and polar systems without the use of additional equipment. Both 2-wire and 4-wire carrier lines may be used.

Carrier frequency assignments are chosen to co-ordinate with existing Western Electric Company type *43A* systems:[1] 18 channels spaced at 170 cps (cycles per second) are obtained between 425 and 3,315 cps. A mark-to-space frequency shift of 70 cycles is used. Eight other channel assignments exist at higher frequencies and these use slightly greater frequency shifts and spacings. The maximum signaling speed of each channel is slightly more than 100 wpm (words per minute) or, in data language, 80 bits per second.

Fig. 1 represents a typical 16-channel terminal showing, from top to bottom, terminal block panel, fuse panels, loop equipment panels, meter panel, jackfield, line equipment panel, and channel equipment shelves. The basic equipment module is four channels.

System Considerations

Fig. 2 is a block diagram of the *23A* system. Each block is self-explanatory, but note that the system is divided functionally into three groups indicated as loop equipment, channel equipment, and line equipment. This same division is also present in the physical arrangements of the terminal equipment, as may be seen by reference to Fig. 1. Advantages of such a separation are twofold. First, the transistorized channel equipment is physically separate from the heat-producing loop equipment. Second, flexibility of design and application is obtained in that the basic channel equipment may be used with many different types of loop and line equipments according to the ultimate system usage. At present, two types of d-c equipment units are designed. One is intended for neutral and polar loop options, and the other for voltage-type hub operation. New applications may well stimulate additional designs, such as might be needed for telemetering converters. The same basic channel unit is also used in the type *2340* subscriber set, a miniature, self-contained carrier telegraph terminal which may be installed in a teleprinter console.

RECEIVE RELAY

A great deal of consideration was given to the use of a transistor as a receive loop keyer in place of the traditional polar telegraph relay. The advantages of using a relay are well known, but the difficulty and expense of maintenance on many common telegraph relays usually outweigh their direct advantages. The polar relay used in the *23A* system (known as the type *PTW 204*), was carefully designed with the emphasis on long service life and extreme ruggedness.[2] Laboratory tests indicate, and user reports verify, that a service life in excess of 1 to 2 years is obtained without readjustment even on full-time high-speed circuits.

The outstanding characteristics of the relay are due primarily to the use of relatively large tungsten-carbide contacts which are fixed to a rigid bar-shaped armature. In use, the contacts slowly powder away without the usual pitting or build-up commonly encountered. Contact burnishing is never required. Field adjustment of the contacts is readily accomplished by resetting each gap to make up for the material which has worn away. There are no other adjustments.

With the maintenance problem under control, full advantage is taken of the relay in design of the loop unit circuitry. Use of a silicon junction diode bridge as an oscillator modulator further simplifies the problem. The basic loop unit may be conditioned for some 40 different loop options including most neutral, polar, effective polar, and inverse neutral types of operation. Selection of the desired option is accomplished by installing a 20-pin option plug which is equipped with the appropriate jumper wiring to select and arrange the various circuit elements. In some of the less commonly used options, the plug contains two or three resistors. Automatic selection of the most appropriate relay contact protection network is also accomplished by the option plug.

The loop unit, then, is seen to contribute great versatility to the system; a flexibility not economically possible with purely electronic keyers such as transistors and vacuum tubes.

CARRIER LINE EQUIPMENT

A minimum requirement of the carrier line equipment panel is balancing and matching transformers for 4-wire lines, and hybrid and matching transformers for 2-wire lines. In addition to such basic facilities, the system design of the *23A* places further requirements on the line equipment.

To permit the use of small inductors in the channel band-pass filters, the output power of an individual channel is −15 decibels referred to 1 milliwatt (−15 dbm). Also, such a level is convenient for group attenuation or amplification to produce the entire range of

---

Paper 59-1111, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted June 10, 1959; made available for printing July 15, 1959.

M. L. STEPHENS is with the Lenkurt Electric Company, Inc., San Carlos, Calif.

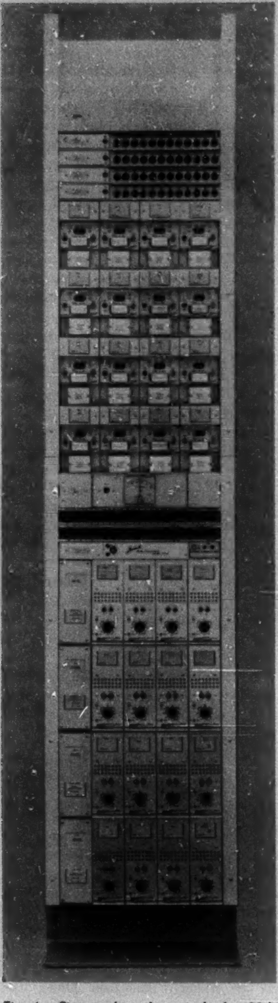


Fig. 1. Sixteen-channel terminal of a 23A system showing channel units, jackfield, meter panel loop equipment units, and fuse panel

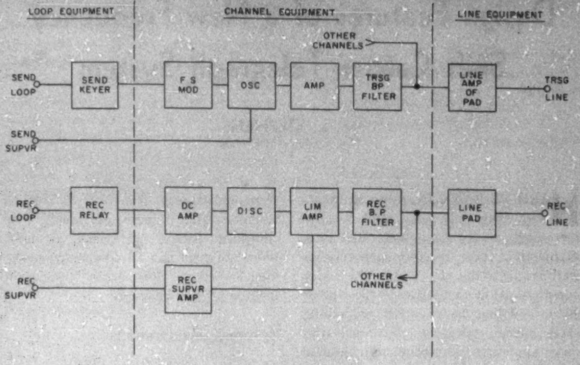


Fig. 2. Block diagram of the 23A system

line power required of the system, which is −40 dbm to +6 dbm. Therefore, the transmit branch of the line equipment panel is provided with an optional adjustable attenuator pad or amplifier as required, and transmit line levels are adjusted on a group or system basis. A similar arrangement is used in the receive branch: an adjustable attenuator pad is provided, and the receiving level is adjusted on a group basis. Such an arrangement has the advantage that all channel units in any system transmit and receive a standard power. Channel units may then be moved from system to system or from spare status to system operation without regard to level adjustments.

Third Function

In addition to mark-and-space functions, the *23A* system provides for low-speed carrier on and off keying. This third function is optional on a channel basis, and is used for certain signaling or supervisory purposes. The channel carrier oscillator is turned on by the application of a ground to the send supervisory lead, and the reception of carrier is indicated by a ground appearance on the receive supervisory lead. Transmission of the third function through a back-to-back channel interconnection is possible without the use of repeating relays.

When equipped, each channel unit is arranged to ground the receive supervisory lead whenever the received channel carrier exceeds a level of approximately −55 dbm. No adjustment of this trip level is necessary and none is provided. Adjustment of the common receive branch attenuator to produce a channel receive level of −40 dbm automatically provides a trip margin of approximately 15 db for every channel in the system. The channel receive filter provides a margin of approximately 8 to 10 db against false operation from noise even if the received carrier to channel noise ratio is as small as 5 to 10 db. Channel noise is defined as that measured in dbm, existing in the band of frequencies from approximately 300 to 3,300 cps.

Circuit Descriptions

Loop Equipment

The function of the loop unit is to modulate the channel carrier oscillator in accordance with the mark-and-space signals generated on the loop by the teleprinter. And, on the receiving side, the unit must key the loop in accordance with the carrier signals received from the distant end. Fig. 3 is a simplified representation of the loop unit arranged for half duplex, using −130-volt telegraph battery, and adjusted for a marking current of 60 ma (milliamperes). With the receive relay contacts closed representing a received mark signal, loop current flows holding the teleprinter selector magnets on mark. The voltage drop through resistor $R1$ is such as to produce a potential of approximately −110 volts at point X. By voltage divider action, the potential at point Y is about −120 volts. The diode quad is therefore held nonconducting by a potential of about 10 volts, and capacitor Cs, the shift capacitor, is isolated from the main oscillator tank circuit by a high impedance and the oscillator produces the high or marking frequency. The diode quad, shift capacitor, and oscillator tank are located with the basic channel unit.

When the teleprinter contacts open to indicate transmission of a space, loop current ceases, point X becomes more negative than point Y, and a current flows through the diode quad. The shift capacitor is therefore coupled into the

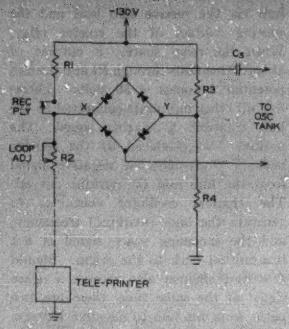

-130V
R1
C5
REC PLY
X
Y
R3
TO OSC TANK
LOOP ADJ
R2
R4
TELE-PRINTER

Fig. 3. Send keyer circuit

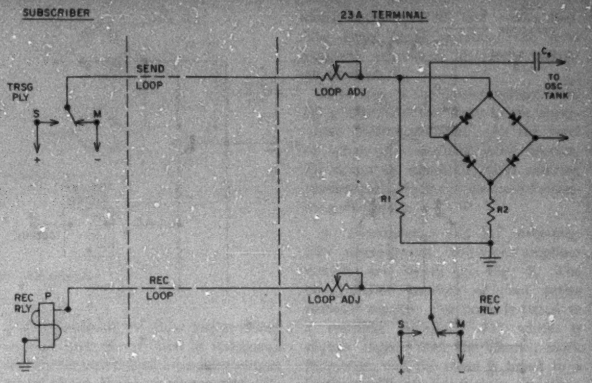

SUBSCRIBER
23A TERMINAL
SEND LOOP
TRSG PLY
S
M
+
-
LOOP ADJ
C5
TO OSC TANK
R1
R2
REC LOOP
REC RLY
P
LOOP ADJ
REC RLY
S
M
+
-

Fig. 4. Two-path polar circuit

tank circuit through the low impedance of the conducting quad, and the oscillator produces the low or spacing frequency.

Now, assume that the teleprinter keyboard contacts are closed, and that the receive relay contacts are open because of the reception of a spacing signal from the distant terminal. Loop current is interrupted by the open relay contacts, point *X* is at ground potential, and point *Y* is still held at about −120 volts by the voltage divider. The diode quad is held nonconducting by a potential of approximately 120 volts, and the oscillator continues to produce the high or marking frequency. Received carrier signals, then, are not transmitted back to the distant terminal.

Different nominal values of loop current are permitted by using a different value of resistance for *R*1, but in all cases *R*1 produces a drop of approximately 20 volts. By changing the value of *R*4, different potentials of telegraph battery may be used. In all cases, the drop across *R*3 is made to be about 10 volts. Positive telegraph battery is accommodated by reversing the connections to the diode quad.

Fig. 4 represents the arrangement used for 2-path polar loops. The voltage developed across *R*1 by the send loop current is used to bias the diode quad. Marking current (from negative battery) biases the quad off: spacing current (from positive battery) biases the quad on. On the receive side, polar signals are developed in the loop directly from the contacts of the receive relay. Selection and arrangement of the proper resistors are done by the jumper wiring of the 20-pin option plug.

The loop unit contains only resistors, capacitors, and the receiving relay; no semiconductors are used (the modulator quad is located within the basic channel unit). The unit is therefore capable of withstanding the temperatures created by the power dissipated by the loop-adjusting rheostats.

Transmission characteristics of the loop unit are similar to those of a system which uses a transmitting relay, in that the oscillator frequency is a function of loop current. Conventional wave-shaping networks may therefore be used whenever indicated.

Hub Equipment

A hub equipment unit is used with the basic channel unit instead of the loop unit when it is desired to operate on a voltage hub basis. In this system, carrier terminals are connected in parallel to a hub lead, and a party-line type of operation results in which transmission received from one channel is transmitted to all other channels. The *23A* hub unit was designed to be compatible with the Western Electric Type *2B* service board where a potential of +60 volts represents marking; −30 volts is spacing; and −60 volts represents a condition known as double space.[3]

Fig. 5 is a simplified schematic diagram of the hub circuitry. The half duplex case is illustrated, in which the

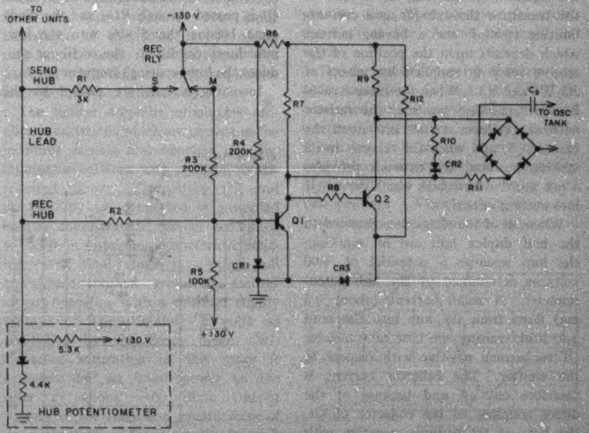

TO OTHER UNITS
-130 V
R6
REC RLY
SEND HUB
R1
3K
S
M
R9
R12
R7
C5
TO OSC TANK
HUB LEAD
R4 200K
R3 200K
R10
CR2
R11
R8
Q2
REC HUB
R2
Q1
R5 100K
CR1
CR3
+130V
5.3K
+130 V
4.4K
HUB POTENTIOMETER

Fig. 5. Electronic hub circuit

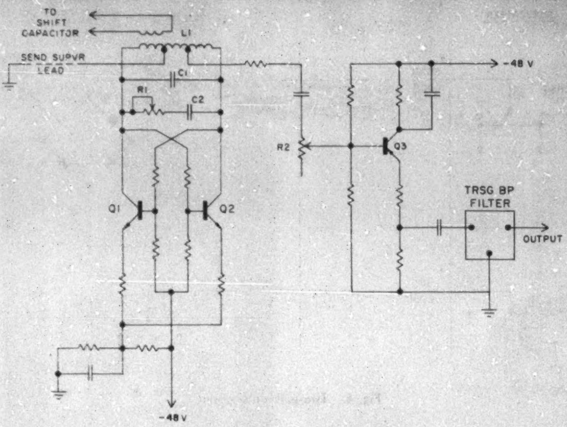


Fig. 6. Oscillator and buffer amplifier

send hub lead and receive hub leads are connected directly together and to a half-duplex hub potentiometer. The hub potentiometer is not a part of the *23A* system, and is located elsewhere. The modulator quad, shift capacitor, and oscillator tank connections are utilized in the same way as in the loop unit, but now the proper voltage and current is supplied by two transistors. *Q*1 and *Q*2 (*2N270*) are connected in tandem in such a way that when *Q*1 is driven to current saturation, *Q*2 is biased to cut off, and vice versa. The base of *Q*1 (input) is subject to bias from two sources: the signals from the hub are applied to the transistor through *R*2 as a current-limiting resistor, and a biasing current which depends upon the position of the receive relay is supplied by means of *R*3, *R*4 and *R*5. When the receive relay is in its marking position, the resistor network produces no net bias upon the base of *Q*1, and when the relay is in its spacing position, the network provides a net positive base bias which drives *Q*1 into current cut-off.

When all of the channels connected to the half duplex hub are on marking, the hub assumes a potential of +60 volts as determined by the hub potentiometer. A small current (about 0.5 ma) flows from the hub into the send hub lead, causing the base of transistor *Q*1 to become positive with respect to the emitter. The collector current is therefore cut off and because of the direct coupling to the collector of *Q*1, the base of *Q*2 becomes negative with respect to its emitter, and its collector current reaches saturation. A portion of the current flows through *R*10 and *CR*2; the resulting voltage drop biases the diode quad into its high impedance condition and the oscillator produces the high (marking) frequency.

When the hub potential is −30 volts, indicating that a spacing signal is to be transmitted, the base of *Q*1 becomes negative with respect to the emitter, and the collector current reaches saturation. Transistor *Q*2 loses its negative base bias, and its collector current becomes cut off. A small reverse bias is provided by *R*12 and *CR*3 to insure full cut-off. A portion of the collector current from *Q*1 is passed through *R*11, and the diode quad biasing the diodes into the low-impedance condition; the oscillator produces the low (spacing) frequency.

Incoming signals are passed to the hub via the receive hub lead and the spacing contact of the receive relay. When the relay spaces, a current of about 30 ma flows through *R*1 and the hub potential becomes −30 volts. Therefore all other hub units connected to the hub transmit a spacing signal. The positive bias produced by the resistor network overcomes the negative signal from the hub and *Q*1 remains cut off. The channel oscillator continues to generate the high (marking) frequency, and the incoming space signal is not transmitted back to the origin. Should a second channel also receive a space signal at the same time, there are two paths from the hub to negative battery, and the hub potential becomes −60 volts. The signal at the base of *Q*1 is now large enough to overcome the positive bias from the resistor network, and a space signal is transmitted back to the origin. In this way, double space signals are always transmitted regardless of the condition of the receive relay. Full duplex hub operation is possible by arranging two hub potentiometers to form an east-west hub and a west-east hub. A jumper connection is made to connect *R*3 to negative battery permanently, and now transmission on the send hub lead is independent of the position of the receive relay, and signals may be passed in both directions at the same time. In either full or half duplex operation, up to ten units may be connected to a single hub.

CHANNEL UNIT

The channel unit contains the frequency-shift transmitting and receiving circuitry required for one channel of a system. For purposes of description, the circuitry is divided into four categories.

1. *Oscillator, Buffer Amplifier, and*

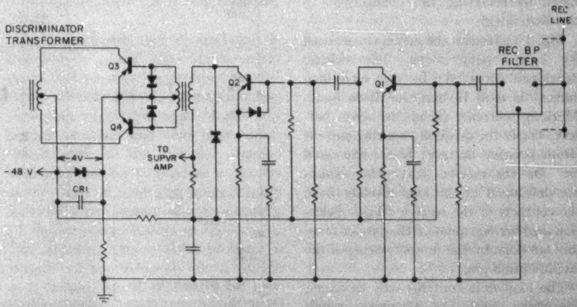


Fig. 7. Limiter amplifier

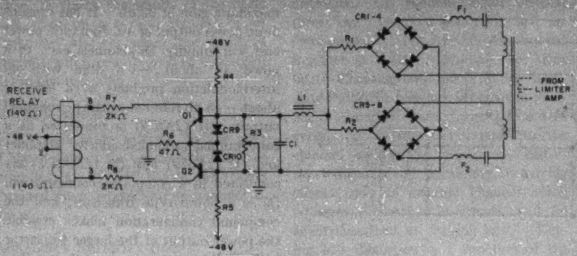


Fig. 8. Discriminator–d-c amplifier

Transmit Filter. Fig. 6 is a schematic diagram of the oscillator, buffer amplifier, and transmit filter. Transistors *Q*1 and *Q*2 (type *2N679*) are the n-p-n type of devices connected in a push-pull oscillator circuit in which the frequency of oscillations is determined by the resonant circuit composed of *L*1, *C*1, and the shift capacitor. *R*1 and *C*2 provide a means of fine frequency adjustment of about ±0.3%. Transistor collector capacitance also appears across the tank circuit, but with the transistors used, capacitance is small compared to the main tank capacitance. Amplitude of oscillations is limited by clipping in the transistors. Oscillations are produced as long as a ground is present on the send supervisory lead; opening of this conductor removes the collector voltage, and oscillations cease.

To obtain good frequency stability as a function of circuit voltage, a ferrite core inductor is used, which is resonated with a polystyrene film capacitor. The combination also gives good stability as a function of temperature because the ferrite core and polystyrene capacitor have nearly equal and opposite temperature coefficients. Any residual coefficient is further reduced by padding of the final tank assembly with capacitors of positive, zero, or negative temperature coefficient as indicated by measurement of the resonant frequency at two ambient temperatures: room temperature and 50 C (degrees centigrade). Coarse frequency adjustment is provided by means of padder capacitors which are brought out to strapping terminals on the outside of the enclosure.

Transistor *Q*3 (type *2N270*) is connected as a common collector amplifier whose primary function is one of impedance transformation from the oscillator to the transmit band-pass filter. Potentiometer *R*2 is provided as a means of precisely adjusting the output to the desired level: a range of approximately 3 db is available, which is just sufficient to take care of variation in transistors and filter insertion loss. Amplifier output power to the filter is approximately −11 dbm.

The transmit band-pass filter provides rejection to the unwanted keying side bands which would otherwise cause adjacent channel interference. The nominal output impedance is 600 ohms, unbalanced, and the out-of-band impedance tends toward open circuit so as to permit flanking (paralleling) of the transmit filters of adjacent channels.

The transmit filter and the oscillator tank are assembled together into a hermetically sealed plug-in case. By changing this one unit, the transmit branch may be operated at any one of the 26-channel frequency assignments.

2. *Limiter Amplifier.* Fig. 7 is a schematic diagram of the limiter amplifier circuits. At the input is shown the receive band-pass filter which serves to select the desired channel frequencies from all those present. The input impedance of the filter is 600 ohms, unbalanced, and it is designed for flanking. Adjacent channel frequencies are rejected by at least 35 db.

The limiter amplifier eliminates amplitude variations which are present in the incoming signals and supplies the discriminator network with a signal of the required power. Transistors *Q*1 and *Q*2 (*2N270*) are connected in grounded emitter configuration; diodes are added to assist in obtaining symmetrical limiting. The final stage is a push-pull transformer-coupled grounded emitter circuit operating with a collector supply voltage of approximately 4 volts as controlled by the Zener diode *CR*1. Collector saturation in this stage is complete for an input power to the filter of about −60 dbm. Output voltage of the stage is a square wave of approximately a 14-volt amplitude, peak to peak, regardless of input level variations between −60 and −15 dbm. The Zener diode maintains the level constant regardless of power supply fluctuations.

A portion of the output voltage from *Q*2 is taken for the input to the supervisory receive amplifier. At the level at which this amplifier must operate, *Q*2 is not in saturation, thus assuring predictable operation of the receive supervisory function.

3. *Discriminator and D-C Amplifier.* The discriminator and d-c amplifier circuits are shown in Fig. 8. The discriminator consists of two series resonant meshes; one mesh is tuned to a resonant frequency *F*1, which is slightly higher than the channel mark frequency, and the other is tuned to a frequency *F*2, which is slightly lower than the channel space frequency. The output of each mesh is individually rectified in full-wave bridge rectifiers and the rectified currents are added in opposition. The resultant difference current flows through the input circuit of the d-c amplifier. Fig. 9 shows the response curve of a typical discriminator network. At the channel mark and space frequencies the output current is approximately ±2 ma respectively. Series resonant circuits are used because they favor the low input impedance (approximately 200 ohms) of the d-c amplifier.

A low-pass filter, *L*1 and *C*1, is included between the discriminator and d-c amplifier. It attenuates carrier frequency energy and other unwanted tones which could cause interference from adjacent channels. Also included, but not shown in Fig. 8, is a reversing switch that permits the proper reception of either the high or low channel frequency for mark.

The d-c amplifier consists of two power-type transistors, *Q*1 and *Q*2 (*2N301A*), connected in a push-pull, grounded emitter configuration. Each transistor drives one coil of the receiving polar relay whose contact position is a function of the difference in current flowing in its two windings. When marking frequency is received, the discriminator mesh tuned

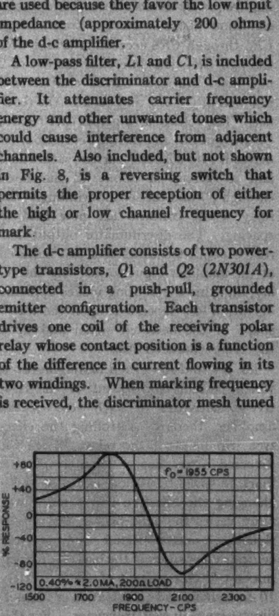


Fig. 9. Typical discriminator response

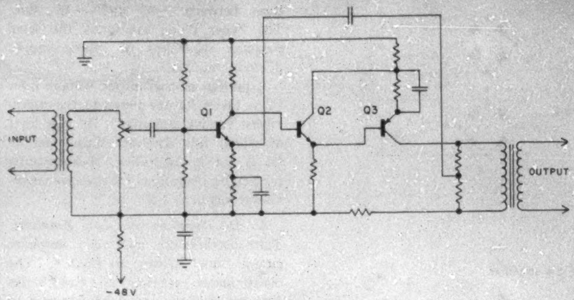


Fig. 10. Group amplifier

to *F*1 has a lower impedance than that tuned to *F*2, and the rectified current from bridge *CR*1 through *CR*4 exceeds that from bridge *CR*5 through *CR*8, and the difference current of 2 ma flows through *L*1, *CR*9, emitter-base junction of *Q*2, and back to the rectifier. The voltage drop across *CR*9 biases *Q*1 into cut-off, and the base current of *Q*2 drives it into collector saturation. A current of approximately 22 ma flows through *R*8 and one winding of the relay and its contacts are held to marking. When spacing frequency is received, the opposite condition obtains: current from bridge *CR*5 through *CR*8 exceeds of bridge *CR*1 through *CR*4, the difference current of 2 ma flows through *CR*10, emitter-base junction of *Q*1, *L*1, and back to the rectifier. The voltage drop across *CR*10 biases *Q*2 into cut-off, and the base current of *Q*1 drives it into collector current saturation, and a current of about 22 ma flows through *R*7 and the second winding of the relay whose contacts are now held to spacing. The discriminator output current is about ten times that just required at the amplifier input to just trip the relay, and this d-c limiting makes dynamic operation less dependent upon the waveshape of the detected signal and upon the relay characteristics.

*R*3, *R*4, and *R*5 form a biasing network which permits adjustment of the relative sensitivities of each side of the amplifier, thereby controlling the effective trip point of the relay. Since the detected signal has a definitely sloping mark to space transition, and vice versa, changing the effective trip point of the relay varies the time bias in the output signals. Under no-signal conditions, the centralized position of the bias control permits a collector current of approximately 10 ma to flow, and a dissipation of 240 milliwatts exists in each transistor. For this reason power-type transistors are used in the amplifier. The small signal transistors used elsewhere in the circuit do not have the necessary dissipation rating.

The entire d-c circuit is truly polar, and variations in relay or transistor parameters which would, in neutral (single-ended) circuits produce bias, are cancelled out.

The receive relay and bias control potentiometer are located in the loop equipment panel along with the other controls and jacks which may be used in day-to-day operation. The discriminator network elements are assembled in the same hermetically sealed plug-in package as the receive band-pass filter. By changing out this one unit, the receive branch may be operated at any of the 26-channel frequency assignments.

4. *Receive Supervisory Circuit.* The received supervisory circuit (not illustrated) consists of a conventional linear grounded emitter a-c amplifier, a voltage doubling detector, a d-c amplifier, and a neutral relay. Gain of the circuit is constant and stabilized with local feedback in such a way that the trip point of the relay is at a receive level of approximately −55 dbm. Stiffness of the circuit is such as to produce an operate-release differential of about 1 db. A time delay capacitor is included to prevent false operation on line hits. The relay contacts are brought out to a terminal block where they may be used for supervisory purposes and, by appropriate strapping, to hold the receive polar relay to mark or space as desired upon loss of carrier.

LINE EQUIPMENT

Of the various elements included within the two basic types of line equipment panels, only the transmit amplifiers will be described in any detail. Fig. 10 is a schematic diagram of the amplifier used for up to 26 channels. The unit is a plug-in, 3-transistor amplifier having a maximum gain of 22 db. It has a maximum power output of +6 dbm total power and will amplify 18 channels each to a power of about −6.5 dbm with all intermodulation products at a level of about −45 db referred to the per-channel output. *Q*1 (*2N679*) is connected as a conventional grounded emitter amplifier and *Q*2 (*2N679*) and *Q*3 (*2N301A*) are connected in a compound configuration. *Q*3 is a power-type transistor, and the compound configuration makes possible the power output of the larger transistor with the linearity and gain of the smaller. Bridge feedback from the output to the input emitter resistor is used to control distortion and output impedance.

When only one channel is to be amplified, as in subscriber set applications, a simpler 2-transistor amplifier is used. The channel amplifier, as it is called, also uses the compound configuration, but the input transistor is omitted, and no overall feedback is provided. Use of the simpler circuit is permissible because, for the same power output an amplifier handling 18 tones must have a peak power output capability of some 12 to 13 db greater than an amplifier for one tone. Single-channel installations, therefore, benefit from the lower cost of the simpler amplifier.

Performance Characteristics

SIGNAL-TO-NOISE CHARACTERISTICS

One of the most important system performance characteristics is its behavior in the presence of randomly distributed (white) noise. For the tests reported here, a method of measurement is used which is rather easily duplicated in the field; signal-to-noise ratio is defined as the ratio of a single-carrier tone power to the rms power of white noise passed through a 3.4-kc low-pass filter. Performance of the system is rated in terms of maximum checkable telegraph distortion in each channel, usually at 100 wpm, and this phrase defines the maximum total distortion (peak distortion) observed over approximately a 5-minute interval.

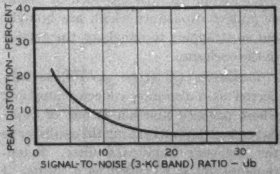


Fig. 11. Peak distortion versus signal-to-noise ratio. Typical channel 100 wpm

Fig. 11 shows the results of the noise tests made on a typical channel of an 18-channel system, using the transmit amplifier and a receive level of −40 dbm at the channel filter.

LIMITING RANGE

Of considerable interest to the user of the system is the permissible range of input levels. Fig. 12 shows the results obtained on a representative channel at 100 wpm.

FREQUENCY STABILITY-BIAS CONSIDERATION

Any frequency shift telegraph system suffers bias distortion from frequency drift in the oscillator, the discriminator, or the transmission facility. If such drifts are small they may be compensated for with the receive bias control, and if they occur slowly with time, they may be ignored between normal line-up intervals. In the *23A* system, a frequency drift of 1 cycle causes a bias distortion of approximately 2% at 100 wpm for the 18 voice-frequency channels. At 60 wpm, the distortion is estimated at about 1.2% per cycle. Slightly less distortion occurs for a given frequency drift for the eight high-frequency channels.

Temperature and voltage fluctions may cause bias distortion in so far as frequency of tuned elements are temperature or voltage dependent. Also, unbalance in the d-c amplifier used to drive the receive relay will cause bias. Through the use of temperature compensation in the tuned networks, a voltage-stabilized limiter circuit, and push-pull d-c amplification, bias distortion is less than 5% at 100 wpm for a 10% change in line voltage or a 15 C change in ambient temperature. Typical channels measure about 2% from either cause.

TOTAL DISTORTION

Other than bias distortion, which may be removed, total channel distortion consists of characteristic and fortuitous distortion. On random traffic signals, these two types of distortion are indistinguishable, and are therefore quoted together as total distortion, other than bias. Characteristic distortion is associated with the response time of the channel to a mark-to-space transition or vice versa, and is therefore related to the channel bandwidth. Fortuitous distortion includes all randomly occurring effects such as internal channel noise, adjacent channel interference, and an effect described as "keying loss." Keying loss distortion is a function of the ratio of channel carrier frequency to the signal or keying frequency; it is minimized by full-wave detection. For a given speed of transmission, keying loss distortion decreases with increasing carrier frequency, and for this reason, the two lowest carrier frequency channels show a greater total distortion than all the others.

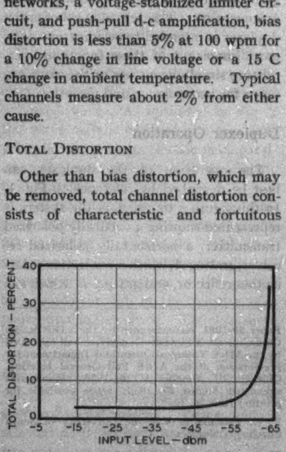


Fig. 12. Typical performance characteristics. Total distortion versus input level at 100 wpm $f_0 = 2{,}805$ cps

For an 18-channel system operating at 100 wpm, the total distortion in any channel is 5% or less, except for the two channels at the low-frequency end of the scale where the distortion is 8% or less. These figures include keying loss and adjacent channel effects.

Equipment Features

TRANSISTOR CONSIDERATIONS

Transistors, as active circuit elements, offer the designer many advantages over vacuum tubes. However, unlike tubes, transistors are quite temperature-sensitive and due care must be taken to prevent the junction temperature from exceeding its rating (usually 85 C) under all operating conditions. In practical terms, power dissipation within the transistor must be derated to account for the equipment temperature rise and ambient temperature expected, and total circuit dissipation must be limited to reduce equipment temperature rise.

The *23A* system is designed for a maximum ambient temperature of 50 C, or 122 F. All transistors are derated for operation to at least 65 C, and temperature measurements on equipment assemblies have indicated a maximum heat rise of about 8 C. Power dissipation within the type *2301A* channel unit is 2.5 watts compared to about 30 watts for comparable vacuum-tube designs. Four different transistor types are used through out the system, and a standardization program currently under way may reduce this number to three.

Fig. 13. The 2301 telegraph channel unit

CHANNEL UNIT

Fig. 13 shows the type *2301A* channel unit. The receive card with the two power transistors is visible. Transmitting, receiving, and supervisory circuits are assembled on separate phenolic cards which are installed in the sheet metal chassis. The unit is plugged into its position in the channel equipment shelf which provides terminal block facilities. The unit face plate provides a card holder for circuit identification, and mounts the switches and test points. An oscillator on-off switch is installed and is used for maintenance purposes, or to turn the oscillator on when the supervisory feature is not used. The second switch

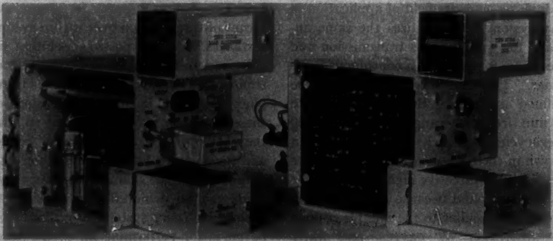


Fig. 14. Type 2332 loop and hub equipment units

is the receive polarity switch. An 11-pin test socket is provided which makes available a number of bridging tests points which are useful in line-up and trouble isolation. The two tuned network assemblies plug into the top of the unit, and can be removed when the channel unit is withdrawn from its mounting. Each circuit card is independently removable from the chassis for servicing, and suitable cable hinges are included to facilitate this operation.

The circuit cards are constructed of phenolic material through which are driven wire staples. Components are installed on the crown side of the board (opposite the legs of the staple), and leads are fed through holes in the board, wrapped around the staple ends, and soldered. Most of the interconnect strapping on the board is done with bare wire, and the number of external cable connections is kept to a minimum.

The mechanical design of the cards is such as to permit much of the assembly to be done by machinery, and automatic fabrication of the cards is gradually being introduced. An adequate description of the process is beyond the scope of this paper.

Other automatic production techniques are being used currently in the production of the filter networks. The process is called "modularized filter construction." Each mesh of the filter consists of a molded plastic module which contains the toroidal inductor and associated capacitor. The modules are tuned (resonated) individually, and then assembled into the final enclosure and connected together to make the complete network. In many cases, padding of the mesh to the required resonant frequency is unnecessary because the toroid has been wound to the precise inductance required on a machine which individually measures the permeability of the core and automatically calculates and winds the correct number of turns.

Loop Equipment Panel

Fig. 14 shows the loop equipment unit (left) and the hub equipment unit. On the back panel of the loop unit is mounted the loop-adjusting rheostats; shaft extensions are used to reach the front panel. The receive relay, option plug, full-duplex half-duplex switch, receive bias control, and closed circuit metering jacks are mounted on the front panel. At the top of the unit is a 24-lug terminal block to which is brought the necessary cabling. Four of the units mount in a shelf, designed to provide sufficient room for cable hinges which permit the units to be withdrawn for inspection without interruption of service.

The card which mounts the transistorized send keyer may be seen on the hub unit in Fig. 13. The front panel of the unit mounts the relay, bias control, and two test points. Hub units mount in the same panel as loop units, but because of the heat generated by the loop unit, the two must not be mounted side by side in a single shelf. Also, hub units must not be mounted directly over loop units. The loop or hub unit is cabled to its associated channel unit with seven conductors.

Conclusions

The *23A* system represents a successful solution to the broad design objectives set forth at the beginning of the paper. Through the use of modern components such as transistors, an improved relay, and modular filters has come a system which is extremely flexible in its application, meets the most rigid performance specifications, and yet is compatible with existing telegraph plant facilities.

References

1. A Carrier Telegraph System for Short Haul Applications. J. L. Hysko, W. T. Rea, L. C. Roberts. *Bell System Technical Journal*, New York, N. Y., vol. XXXI, July 1952, no. 4, pp. 666–87.

2. An Improved Polar Telegraph Relay, W. D. Cannon, T. Rystedt. *Electrical Engineering*, vol. 70, no. 12, Dec. 1951, pp. 1088–92.

3. A New Telegraph Serviceboard Using Electronic Circuits, J. R. Davey, F. H. Hanley, M. R. Purvis. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, Mar. 1955, pp. 30–37.

Vol.
Pt. I
195

A Circular-Polarization Duplexer for Millimeter Waves

R. G. FELLERS
MEMBER AIEE

THIS PAPER describes a circular-polarization duplexer for use in the millimeter (mm) wavelength range. Such a duplexer finds its field of application in any system which utilizes the same antenna for simultaneous transmission and reception, for example, a continuous-wave radar. The mm wavelengths are here defined as covering the range from 1 to 10 mm (frequency range 30,000 to 300,000 megacycles).[1] Much interest has been shown in these frequencies because of the possibility of directive high-gain antennas of small size and because of atmospheric properties peculiar to the mm range such as absorption by oxygen and water vapor.[2–4]

The fact that dominant (TE_{10}) mode rectangular waveguides possess extremely high attenuation in this frequency range has led to a search for alternative means of transmission over within-system distances.[5–8] One possible solution to this intrasystem problem is the use of transmission between two antennas to replace a waveguide run.[9] Dimensions of these antennas may be selected so that they are operating in the Fresnel region and transmission loss much less than that obtainable in the dominant mode waveguide can be achieved. In considering such a free-space system, provision must be made for performing all of the functions which are performed by the waveguide elements in the more conventional arrangement. The circular-polarization duplexer is intended to perform one of the most important of these functions: permitting simultaneous use of the same antenna for transmission and reception.

Duplexer Operation

The operation of the duplexer can best be described by reference to Fig. 1. In this figure a top view of the system is represented showing a vertically polarized transmitter, a horizontally polarized receiver, grating *A*, which is a selective polarization reflector, and grating *B*, which is a

---

Paper 59-1085, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted August 29, 1958; made available for printing June 24, 1959.

R. G. Fellers is with the University of South Carolina, Columbia, S. C.

The research upon which this paper was based was supported by the U. S. Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under contract no. *AF-18(603)-43*.

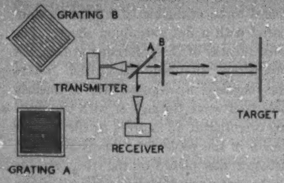


Fig. 1. Duplexer system diagram

device for producing circular polarization. The enlarged views of gratings *A* and *B* at the left of the figure represent the gratings in the normal vertical position as seen from the transmitting side. A vertically polarized wave comes from the transmitter and passes through grating *A* which is so oriented that there is virtually perfect transmission. The wave then passes through grating *B* and emerges circularly polarized. Upon reflection from the target the sense of the circular polarization is reversed and after passing back through grating *B* the wave emerges as a linear horizontally polarized wave which is completely reflected by grating *A* into the receiving horn.

The selective reflector, grating *A*, consists of closely spaced metal slats so oriented that a vertically polarized wave has its electric field normal to the slats and hence passes through with extremely small reflection. The spacing between slats is such that each pair behaves as a waveguide which is well below cutoff at the operating frequency for waves polarized parallel to the slats. As a result it acts as an almost perfect reflector for horizontally polarized waves.

Grating *B* consists of slats which are oriented at 45 degrees with respect to the electric vector of the incoming vertically polarized wave. As indicated in Fig. 2 this wave may be resolved into two components, one with its electric field parallel to the slats and one with its electric field normal to the slats.

The magnitude of each of the components of the wave is 0.707 times the magnitude of the incident field. The component normal to the slats is virtually unchanged by the presence of the grating and passes through at the velocity of a plane wave in free space. The slats serve as a group of parallel waveguides for the field polarized parallel to the slats and this component propagates at the phase velocity for a waveguide of width corresponding to slat spacing. This spacing is selected as 0.280 inch for the operating free-space wavelength of 8.5 mm. The guide wavelength in a guide of this dimension is then 10.61 mm as computed from the well-known relation:

$$\lambda_g=\frac{\lambda}{\sqrt{1-(\lambda/2a)^2}}$$

where

λ = free-space wavelength = 8.5 mm
λ_g = guide wavelength = 10.61 mm
a = slat spacing = 0.280 inch = 7.115 mm

$$\lambda_g=\frac{8.5}{\sqrt{1-(8.5/14.230)^2}}=\frac{8.5}{0.801}$$
$$=10.61 \text{ mm}$$

The dimension of the grating in the direction of propagation, which corresponds to the waveguide length, was selected as 10.61 mm which is exactly one guide wavelength and 10.61/8.5 = 1.25 free-space wavelengths. With these dimensions the component parallel to the slats emerges from the grating advanced in phase by 1/4 free-space wavelength or 90 degrees in phase with respect to the component normal to the slats.

This grating has then produced two components oriented perpendicular to each other and 90 degrees out of time phase which constitutes a circularly polarized wave as described by the equations:

$$E_x=0.707E_0\cos(\omega t-\beta z+90°)$$

$$E_y=0.707E_0\cos(\omega t-\beta z)$$

with reference to the co-ordinate system of Fig. 2, the positive *z*-axis being directed into the paper. The electric vector in this circularly polarized wave may be regarded as rotating in the clockwise direction as viewed looking in the direction of propagation

Upon reflection from the plane, almost perfectly conducting target, the expressions for the field components become:

$$E_x=-0.707E_0\cos(\omega t+\beta z+90°)$$

and

$$E_y=-0.707E_0\cos(\omega t+\beta z)$$

This represents a wave propagating in the negative *z*-direction with the electric vector rotating counter clockwise as viewed looking in the direction of propagation which is now away from the target.

When this wave again reaches the polarizing grating, the component of field along the *x*-axis is still parallel to the grating slats and, on passing through the grating, is advanced by an additional 90 degrees with respect to the component of field along the *y*-axis.

After passing through the grating the components are represented by the equations:

$$E_x=-0.707E_0\cos(\omega t+\beta z+180°)$$
$$=0.707E_0\cos(\omega t+\beta z)$$

and

$$E_y=-0.707E_0\cos(\omega t+\beta z)$$

In these equations β is the propagation constant measured in radians per meter and ω is the angular frequency measured in radians per second.

These components can now be recombined to represent a horizontally polarized wave as indicated in Fig. 2. This wave is then completely reflected from grating *A* into the receiver which is polarized to receive only horizontally polarized waves.

This device accomplishes the duplexing operation with no loss in power due to the action of the duplexer itself. The field components are recombined to yield a horizontally polarized field, equal in strength to the transmitted vertically polarized field, except, of course, for the decrease in amplitude resulting from traveling to the target and back which is inherent in any system of this sort.

Experimental Results

Experimental evaluation of the duplexer system was carried out with the arrangement indicated in Fig. 1 using the apparatus shown in Fig. 3.

All experimental work was done at a frequency of 35,300 megacycles which corresponds to a wavelength of about 8.5 mm. A Raytheon *QK 291* reflex klystron was employed as the radio-frequency oscillator with an appropriate power supply, square-wave-modulated at 1,000 cycles. The detectors used were type *1N53* crystals in properly designed mounts. The square-wave output of these detectors was amplified by low-noise amplifiers tuned to 1,000 cycles and dis-

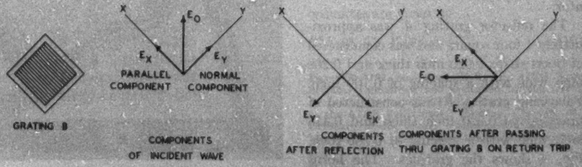

GRATING B
X
E$_O$
Y
E$_X$
E$_Y$
PARALLEL COMPONENT
NORMAL COMPONENT
COMPONENTS OF INCIDENT WAVE
X
Y
E$_Y$
E$_X$
COMPONENTS AFTER REFLECTION
X
Y
E$_X$
E$_O$
E$_Y$
COMPONENTS AFTER PASSING THRU GRATING B ON RETURN TRIP

Fig. 2. Diagram of grating B

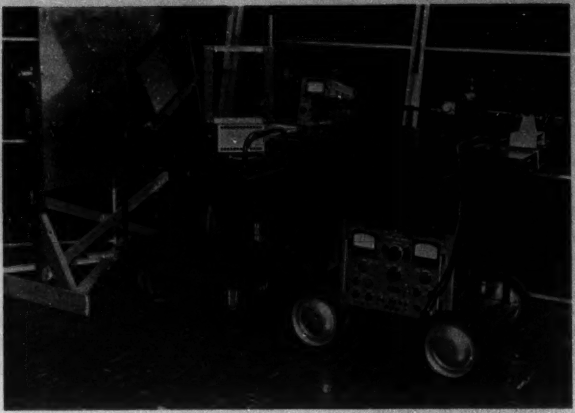


Fig. 3. Apparatus for duplexer system

played on a conventional meter after rectification of the amplified 1,000-cycle signal.

All relative power measurements were made with a Polytechnic Research and Development Company calibrated attenuator type *192F1* with a presumed accuracy of 0.2 db (decibel). In order to eliminate crystal law and amplifier nonlinearity, power levels were adjusted to the same value by adjusting the attenuator and differences in level were obtained from the attenuator calibration. For example, in measuring transmission loss, the crystal detector was connected to the transmitter and the attenuator set at some high value of attenuation to produce a convenient meter reading on the output of the amplifier. The detector was then moved to the receiver and the attenuator readjusted to yield the same meter reading as that obtained when it was connected to the transmitter. The difference in attenuator settings then provides the value of the transmission loss and all system inaccuracies except those in the attenuator are eliminated. Output power from the oscillator was maintained constant by use of a directional coupler power monitor and a second adjustable attenuator.

The reflector grating *A* was approximately 1 foot square and was constructed of brass strips 0.020 inch thick and 0.200 inch wide with a spacing of 0.100 inch. Polarizing grating *B* was constructed of brass strips 0.020 inch thick and 0.418 inch wide with a spacing of 0.280 inch. Fig. 4 shows the gratings. The horns used for transmission and reception had aperture dimensions of 2.925 by 2.132 inches and were 25.6 inches long. A sheet of aluminum sufficiently large to intercept virtually all of the antenna beam was used as a target. Reflections from floor and surrounding objects were eliminated by properly disposed absorptive material.

As a prelude to evaluating the performance of the duplexer system as a whole, essential characteristics of various components were measured.

Transmission loss through the reflector grating *A* was measured with the electric field polarized normal to the grating slats and was found to be negligible (<0.1 db). Loss on reflection from this grating was measured with the electric field polarized parallel to the grating slats and was found to be negligible (<0.1 db) as compared with a flat aluminum sheet. The effectiveness of the polarizing grating for producing circularly polarized waves was checked by transmitting a linearly polarized wave through it with the grating oriented at 45 degrees to the electric vector in the wave. A linearly polarized detector was placed on the receiving side of the grating and relative received power level was measured as function of angle between the polarization of the detector and the vertical. For full 360-degree rotation of the detector, the variation in received signal was less than 0.3 db from maximum to minimum, indicating very good achievement of circular polarization.

The duplexing system was then arranged in operating position as shown in Fig. 1 and the target was replaced by absorptive material to permit measurement of transmitter power which reaches the receiver due to imperfections in the duplexer. Isolation in the original setup was found to be about 30 db. A matching device which was designed and attached to the grating increased this figure to about 50 db.

Experimental operation of the duplexer system was then carried out. In view of the characteristics of the component parts no difficulty was expected and none was encountered. The duplexer operated as predicted with the aluminum plate as a target and very little loss (≤ 0.5 db) due to the duplexer was detected. This was ascertained by comparison with measurements taken using transmitter and receiver horns side by side at the same distance from the target as that used in the duplexer operation and by direct transmission measurements between transmitter and receiver horns at twice this distance.

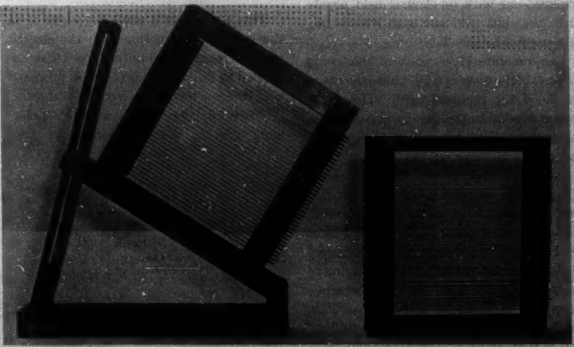


Fig. 4. Reflector and polarizing gratings

Conclusions

From the foregoing results it is clear that a circular-polarization duplexer designed according to the principles of this paper functions as predicted with very small duplexer loss. This is an advantage as compared with the magic-T or hybrid-junction continuous-wave duplexer which has an inherent 6-db loss.

It might be contended that a plane target is very improbable and that the return signal would not be circularly polarized and that this would result in some loss. This is a valid criticism but one which is also applicable to linearly polarized systems. Any target may have characteristics which yield reflected signals of different polarization than those incident upon it. To explore this question fully, it would be necessary to accumulate extensive data on the reflection properties of typical targets for incident radiation of various polarizations.

A free-space mm wave transmission system including a duplexer as described is most likely to find application in the use of extremely short wavelengths (1 to 3 mm) where the physical size of the components is small. It can be used in a radar system as described here or in a communications system in which both transmitting and receiving antennas produce circular polarization.

References

1. Millimeter Waves and Their Applications, R. G. Fellers. *Electrical Engineering*, vol. 75, no. 10, Oct. 1956, pp. 914–17.

2. Further Theoretical Investigations on the Atmospheric Absorption of Microwaves, J. H. Van Vleck. *Report no. 664*, Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., Jan. 1945.

3. Absorption of Microwaves by Oxygen, J. H. Van Vleck. *Physical Review*, New York, N. Y., vol. 71, Apr. 1947, pp. 413 ff.

4. Propagation of Short Radio Waves (book), Kerr. McGraw-Hill Book Company, Inc., New York, N. Y., 1951, chap. 8.

5. Low-Loss Waveguide Transmission, S. E. Miller, A. C. Beck. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, Mar. 1953, pp. 348–58.

6. An Investigation of Dielectric Rods as Waveguides, C. H. Chandler. *Journal of Applied Physics*, New York, N. Y., vol. 20, Dec. 1949, p. 1188.

7. Attenuation in Dielectric Circular Rod, W. M. Elsasser. *Ibid.*, p. 1193.

8. Dielectric Waveguide Techniques for Millimeter Waves, A. G. Fox. *Report No. MM-51-160-20*, Bell Telephone Laboratories, New York, N. Y., June 4, 1951.

9. Non-Waveguide Transmission of Millimeter Waves, R. G. Fellers, W. E. Propst. *ASTIA Document no. AD126 483*, University of South Carolina, Columbia, S. C., Apr. 15, 1957.

Conducting Sphere in Alternating Magnetic Fields

H. PORITSKY
MEMBER AIEE

THE FIELD DISTURBANCE produced by a conducting metallic body in an alternating magnetic field is of interest from several points of view, for instance in connection with eddy-current losses, induction heating, and in high-speed rotation of a sphere induced by a revolving field. Again, the location of a hidden metallic body, a problem met in a number of industries, is often carried out by producing an alternating magnetic field and detecting the presence of the body from the field distrubance produced by it.

In this paper an analysis is carried out of the field disturbance produced by a conducting sphere placed in a uniform alternating magnetic field. The electromagnetic oscillations of a dielectric sphere have been discussed by J. J. Thompson.[1] The following solution was obtained by the author in an unpublished paper in 1948, and has since appeared in the literature.[2] For this reason, its derivation will be reviewed very briefly.

By using a complex permeability, $\mu=\mu_0\epsilon^{-j\delta}$, one may take approximate account of magnetic hysteresis.

The disturbance produced by an oblong body of rotational symmetry is also briefly discussed.

Field Solution

The conducting sphere of radius a, permeability μ, and resistivity ρ in ohm-cm (centimeter), is placed so that its center is at the origin, and axes chosen so that the z-axis is parallel to the applied, uniform alternative field, of frequency f. In addition to the rectangular co-ordinates and their unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, also introduced are cylindrical co-ordinates r, φ, z with unit vectors $\mathbf{i}$, $\mathbf{i}_\varphi$, $\mathbf{k}$ in directions of increasing r, φ, z, and spherical co-ordinates R, θ, φ, and vectors $\mathbf{i}_R$, $\mathbf{i}_\theta$, $\mathbf{i}_\varphi$, where

$$x=r\cos\varphi,\quad y=r\sin\varphi \tag{1}$$

$$\cos\theta=z/R,\quad R^2=r^2+z^2 \tag{2}$$

Time is assumed to enter as a factor

$$\epsilon^{j\omega t},\ \omega=2\pi f,\ j=\sqrt{-1} \tag{3}$$

where f is the frequency, but generally this factor is omitted.

Upon neglecting displacement currents, Maxwell's equations become, in practical centimeter-gram-second units

$$\nabla\times\mathbf{H}=0.4\pi\mathbf{I}=0.4\pi\sigma\mathbf{E}$$

$$\nabla\times\mathbf{E}=-10^{-8}j\omega\mu\mathbf{H}$$

$$\nabla\cdot(\mu\mathbf{H})=0,\ \Delta\cdot(\mathbf{E})=0 \tag{4}$$

where $\mathbf{E}$ is in volts/cm, $\mathbf{I}$ the current density vector, and $\sigma(=1/\rho)$ the conductivity. Elimination shows that the $\mathbf{E}$, $\mathbf{H}$ vectors, as well as their scalar Cartesian components, satisfy the differential equation

$$\nabla^2u=\alpha^2u \tag{5}$$

where

$$\alpha^2=4\pi\omega\sigma\mu\,10^{-9}j,\ \alpha=(1+j)\beta,\qquad \beta=2\pi\sqrt{f\sigma\mu\,10^{-9}} \tag{6}$$

From equation 5, it is evident that α (and hence also β) are of dimensions 1/length; $1/\beta$ is called the "depth of penetration" of the material at the frequency in question, since the solution of equation 5

$$u=\epsilon^{-\alpha y}=\epsilon^{-j\beta y}\,\epsilon^{-\beta y} \tag{7}$$

shows that a uniform plane parallel (magnetic) field penetrating a semi-infinite solid $y>0$, decays as $\epsilon^{-\beta y}$.

In free space $\mu=1$, $\sigma=0$. The first Maxwell equation 4 yields

$$\nabla\times\mathbf{H}=0 \tag{8}$$

while equation 5 reduced to the Laplace equation

$$\nabla^2u=0 \tag{9}$$

Solutions of equation 9 in spherical co-ordinates are given by

$$R^nY_n(\theta,\varphi),\ Y_n(\theta,\varphi)/R^{n+1} \tag{10}$$

Paper 59-1102, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted November 5, 1958; made available for printing July 3, 1959.

H. Poritsky is with the General Electric Company, Schenectady, N. Y.

where R^nY_n is a homogeneous polynomial of degree n in x, y, z, and Y_n is a "surface spherical harmonic" of degree n; it can be represented as a linear combination of

$$(\cos m\varphi,\ \sin m\varphi)P_n^{(m)}(\cos\theta),\quad m\leqq n \tag{11}$$

where $P_n^{(m)}$ are the "associated" Legendre functions.

Inside the conducting sphere, similar pertinent solutions of equation 5 are given by

$$Y_nS_n(\rho),\quad \rho=\alpha R \tag{12}$$

where S_n satisfy the differential equation

$$\frac{d^2S_n}{dR^2}+\frac{R}{2}\frac{dS_n}{dR}-\frac{n(n+1)}{R^2}S_n+\alpha^2S_n=0 \tag{13}$$

Solutions of equation 13 for $n=0, 1, \ldots, n$, which are analytic at $R=0$ are given by[1]

$$\left.\begin{aligned} S_0&=\frac{\sinh\rho}{\rho}\\ &\vdots\\ S_n&=\rho^n\left(\frac{1}{\rho}\frac{d}{d\rho}\right)^n\left(\frac{\sinh\rho}{\rho}\right)\end{aligned}\right\}\rho=\alpha R \tag{14}$$

and S_n may be also expressed in the form

$$R^{-1/2}J_{n+(1/2)}(\rho) \tag{15}$$

since equation 13 may be converted to a Bessel differential equation of order $n+(1/2)$ for the function $R^{1/2}S_n$.

Field Solution for Sphere

Let the applied alternating uniform magnetic field pointing in the z-direction be

$$\mathbf{H}=H_0\mathbf{k} \tag{16}$$

(The exponential factor in equation 3 is implied.) It is evident that the complete field will be symmetric about the z-axis. Indeed, integration of the second Maxwell equation 4 for this $\mathbf{H}$ yields as a possible electric field

$$E=(Cy\mathbf{i}-Cx\mathbf{j})=-Cr\mathbf{i}_\varphi,\quad C=j\omega10^{-8}H_0/2 \tag{17}$$

with the electric lines of force running in circles about the z-axis.

The scalar x, y components in equation 17 are of the form R^nY_n with $n=1$, and

$$Y_1=\left(\frac{y}{R},\frac{-x}{R}\right)=(\sin\theta\sin\varphi,\ -\sin\theta\cos\varphi) \tag{18}$$

It turns out possible to complete the field solution including the eddy currents, with the same Y_1 by putting

$$\mathbf{E}=C\left(\frac{y}{R},\frac{-x}{R},0\right)\left(R+\frac{A}{R^2}\right),\quad R\geqq a \tag{19}$$

outside the sphere, and

$$\mathbf{E}=C\left(\frac{y}{R},\frac{-x}{R},0\right)BS_1(\rho),\quad R\leqq a \tag{20}$$

inside the sphere.

These equations imply that in cylindrical components $\mathbf{E}$ reduces to $E_\varphi\mathbf{i}_\varphi$ where E_φ is independent of φ

$$\mathbf{E}=E_\varphi\mathbf{i}_\varphi=-C\sin\theta\left(R+\frac{A}{R^2}\right)\mathbf{i}_\varphi,\quad R\geqq a \tag{21}$$

$$\mathbf{E}=\mathbf{E}_\varphi\mathbf{i}_\varphi=-C\sin\theta\,BS_1(\alpha R)\mathbf{i}_\varphi,\quad R\leqq a \tag{22}$$

The constants A and B are now to be determined so that the tangential components of $\mathbf{H}$ and $\mathbf{E}$ are continuous at the spherical boundary $R=a$.

Making the magnitude E_φ continuous at $R=a$, one obtains

$$a+\frac{A}{a^2}=BS_1(\alpha a) \tag{23}$$

Upon making H_θ along $R=a$ continuous, one is led to

$$\frac{d}{dR}\left(R^2+\frac{A}{R}\right)=\frac{1}{\mu}\frac{d}{dR}BRS_1(\alpha R)\quad\text{for } R=a \tag{24}$$

Solving for A there results

$$A=a^3\frac{(2\mu-1)-F(\alpha a)}{(\mu+1)+F(\alpha a)} \tag{25}$$

where

$$F(\alpha a)=\frac{(\alpha a)S_1'(\alpha a)}{S_1(\alpha a)}=\frac{-2(\alpha a)+[(\alpha a)^2+2]\tanh(\alpha a)}{(\alpha a)-\tanh(\alpha a)} \tag{26}$$

From the second Maxwell equation 4 and equation 19, one may solve for $\mathbf{H}$. For $R>a$ it turns out that the resulting magnetic field is derivable from a scalar potential V

$$\mathbf{H}=\Delta V,\quad R\geqq a \tag{27}$$

where

$$V=H_0\left(z-\frac{Az}{2R^3}\right)=H_0P_1(\cos\theta)\left(R-\frac{A}{2R^2}\right) \tag{28}$$

Indeed, the existence of a scalar potential for $\mathbf{H}$ for $R>a$ follows from equation 8.

It can be verified that for all R, the magnetic field $\mathbf{H}$ also possesses a flux function f which is axially symmetric and such that

$$\mu\mathbf{H}r=\nabla'f \tag{29}$$

where

$$\nabla'=\mathbf{i}_r\frac{\partial}{\partial z}-\mathbf{k}\frac{\partial}{\partial r}=\mathbf{i}_\theta\frac{\partial}{\partial R}-\mathbf{i}_R\frac{\partial}{R\partial\theta}$$

and f is given by

$$f=j(10^8/\omega)rE_\varphi \tag{30}$$

This can also be inferred from the emf (electromotive-force) law, applied to the circles of constant r and z, since

$$\psi=2\pi f \tag{31}$$

is the magnetic flux linking them.

The potential

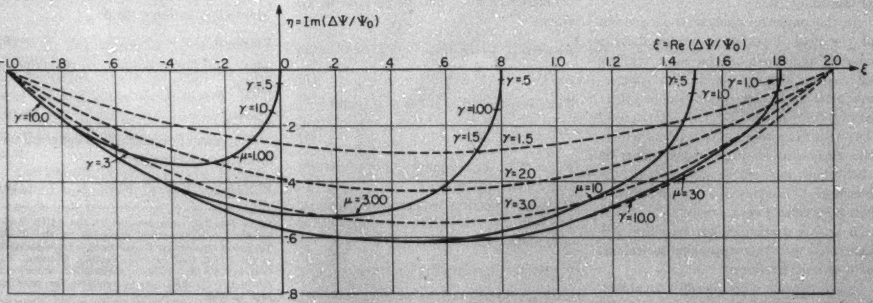


Fig. 1. Plot of $\Delta\psi/\psi$: for various μ, γ

Table I. Values of $\Delta\psi/\psi_0$

γ	μ = 1	3	10	30	100	1,000
0.01				$1.9032+0j$		$1.9940-0.0^6120$
0.03				$1.9032+0j$		$1.9940-0.0^5108j$
0.3	$-0.0^3458\ -0.01258j$	$0.7996-0.01350j$	$1.4998-0.0079j$	$1.8124-0.0038j$	$1.9423-0.0006j$	$1.9940-0.0^3109j$
0.5	$-0.0^22600-0.03317j$	$0.7777-0.03589j$	$1.4989-0.0208j$	$1.812\ -0.0088j$	$1.9423-0.0015j$	$1.9940-0.0^3300j$
1.0	$-0.02541\ -0.1316j$	$0.7800-0.1447j$	$1.4923-0.0846j$	$1.811\ -0.0359j$	$1.9423-0.0061j$	$1.9939-0.0^2119$
1.5	$-0.1060\ -0.2520j$	$0.7103-0.2938j$	$1.4641-0.1786j$	$1.800\ -0.0766j$	$1.9386-0.0130j$	$1.9935-0.0^2287$
2.0	$-0.2457\ -0.3343j$	$0.5702-0.4344j$	$1.3999-0.2869j$	$1.780\ -0.1268j$	$1.9312-0.0219j$	$1.9930-0.0^2446j$
3.0	$-0.5475\ -0.3014j$	$0.1233-0.5212j$	$1.1007-0.4727j$	$1.656\ -0.2419j$	$1.9044-0.0451j$	$1.9857-0.0^2781j$
10.0	$-0.8500\ -0.1350j$	$-0.5637-0.3325j$	$0.2009-0.5956j$	$1.117\ -0.5255j$	$1.7049-0.1434j$	$1.9699-0.0298j$

$\gamma=1$, $\mu=10\ (1-3j)$, $\Delta\psi/\psi_0=1.51530-0.196443j$.

$$-2H_0A\ \frac{z}{R^3} \tag{32}$$

which is added to H_0z, the potential of the initial applied field, will be recognized as the potential of a magnetic dipole of moment

$$M=-2AH_0 \tag{33}$$

placed at the origin, with its axis in the z-direction. It is easy to verify that the same net external field results if the sphere is replaced by a (nonconducting) sphere of radius a, magnetized in the z-direction, to a uniform flux density

$$B_z=H_0\left(1+\frac{4A}{a^3}\right) \tag{34}$$

Flux Increase Over Equatorial Plane

The increase in flux $\Delta\psi$ over the equatorial plane $r=a$, $z=0$ of the sphere, as a fraction of the initial flux ψ_0 passing through the equator in absence of the sphere will now be calculated.

Equations 21, 25, 30, 31 yield

$$\frac{\Delta\psi}{\psi_0}=\frac{\Delta E}{E_0}=\frac{A}{a^3}=\frac{2\mu-1-F(\alpha a)}{\mu+1+F(\alpha a)}=\xi+j\eta \tag{35}$$

where ξ is the real part, η the imaginary part of $\Delta\psi/\psi_0$.

As a limiting case consider $\omega\rightarrow\infty$. Now $|\alpha a|$ is large numerically, tanh $(\alpha a)=1$, thus reducing $F(\alpha a)$ to $|\alpha a|$ and $\Delta\psi/\psi_0$ to -1. This means that the increase in flux is -100%, corresponding to zero net flux crossing the region occupied by the sphere. This is due to the pronounced skin effect which keeps the flux away from the interior of the sphere.

It will be noted that large μ keeps the limit -1 from being reached even when $|\alpha a|$ is large, until $|\alpha a|$ is larger than μ. Apparently, even with a pronounced skin effect and thin penetration, a fair amount of flux may crowd into the skin layer if its permeability is large.

As a further limiting case, consider the other extreme of very complete flux penetration, or small $|\alpha a|$. We now obtain from equations 26 and 35, on expanding the hyperbolic functions in powers of $\alpha a=\rho$,

$$F(\alpha a)=\frac{\rho^2\left(\frac{1}{3}\right)+\dots}{\rho^2\left(\frac{1}{3}\right)+\dots}=1+\dots \tag{36}$$

and equation 35 yields

$$\frac{\Delta\psi}{\psi_0}=\frac{2\mu-2}{\mu+2} \tag{37}$$

This agrees with the flux increase for the d-c field case. In particular, for $\mu=\infty$ there results $\Delta\psi/\psi_0=2$, leading to the well-known result that the net flux through the sphere is three times the flux crossing the space occupied by the sphere previous to its introduction.

As a numerical example consider

$$\mu=1$$

$$1/\sigma=60\ 10^{-6}\ \text{ohm-cm}$$

$$a=1\ \text{cm}$$

$$\alpha a=(1+j)$$

Since $a=1$, $\alpha=1+j$, $\alpha^2=2j$ there follows

$$\alpha^2=j4\pi\mu\omega\sigma10^{-9}=2j$$

$$\omega=\frac{2}{4\pi\sigma10^{-9}}$$

$$=\frac{2(60)10^{-6}}{12.6\ 10^{-9}}=9{,}530$$

this corresponds to a frequency of 1,516.

The calculation of the disturbance in the field proceeds as follows

$$\tanh\,(\alpha a)=\tanh\,(1+j)$$

$$=\frac{\tanh 1.0+j\tan 1.0}{1+j\tanh 1.0\tan 1.0}$$

$$=\frac{0.7616+j1.5574}{1+j0.7616(1.5574)}$$

$$=\frac{0.7616+j1.5574)(1-j1.187)}{1+1.408}$$

$$=1.084+j0.271$$

$$F(\alpha a)=\frac{(a\alpha)S_1'}{S_1}=\frac{((\alpha a)^2+2)\tan(\alpha a)-2(\alpha)a}{\alpha a-\tanh\,(\alpha a)}$$

$$=\frac{(2j+2)(1.084+j0.271)-2(1+j)}{(1+j)-(1.084+j0.271)}$$

$$=\frac{2(1+j)(0.084+j0.271)}{-0.084+j0.729}=1.018+j0.396$$

Substitution in equation 35 yields

$$\frac{\Delta\psi}{\psi_0}=\frac{(2-1)-F(\alpha a)}{1+1+F(\alpha a)}=\frac{1-1.018-j0.396}{2+1.018+j0.396}$$

$$=\frac{0.018+j0.396}{3.018+j0.396}$$

$$=-\frac{(0.018+j0.396)(3.018-j0.396)}{9.12+0.1568}$$

$$=-0.0227-j0.128$$

Calculations of $\xi+j\eta=\Delta\psi/\psi_0$ were carried out for various values of γ in

$$(\alpha a)=(1+j)\gamma,\ \gamma=\beta a \tag{38}$$

and various values of μ. The results are shown on Fig. 1 in the complex $(\xi+j\eta)$-plane. As will be noted, the points are plotted for fixed values of μ and varying values of γ, and for fixed values of γ and varying values of μ. It will be noted from equation 35 that for fixed γ, and hence fixed αa, the variable $\xi+j\eta$ is linear rational in μ. Hence for fixed γ, as μ varies, the locus of $\xi+j\eta$ may be shown to describe an arc of a circle passing through the points $\xi+j\eta=-1$ (for $\mu=0$) and $\xi+j\eta=2$ (for $\mu=\infty$).

The values of $\Delta\psi/\psi_0$ are also displayed in Table I.

Detection of Sphere by Means of Circular Coils

Consider a circular coil of radius b, placed as in Fig. 2, in the plane $z=d$, so that its center is $(0, 0, d)$. The fractional change in the induced emf in the coil, due to the introduction of the sphere, may be found from the fractional increase in the flux linkage of the coil, $\Delta\psi_1/\psi_1$. Equations 21, 30, 31, 35 yield

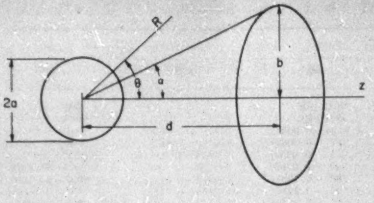


Fig. 2 (left). Sphere and detecting coil

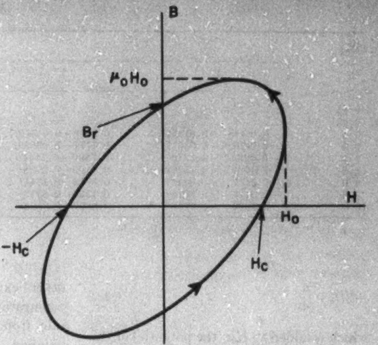


Fig. 3 (right). Elliptic hysteresis loop

$$\Delta\psi_l=\pi H_0\frac{a^3b^2(\xi+j\eta)}{(b^2+d^2)^{3/2}} \tag{39}$$

and hence

$$\frac{\Delta\psi_l}{\psi_l}=\frac{a^3(\xi+j\eta)}{(d^2+b^2)^{3/2}} \tag{40}$$

Suppose next that the coil axis, while passing through the sphere center, makes an angle ζ with the initial field direction. The initial flux linked by the coil is evidently reduced by a factor cos ζ. It may be shown that the change in the flux (linkage) due to the sphere is reduced in the same ratio. Hence equation 40 still applies.

The magnetic field of the dipole of the moment M given by equation 33 can also be produced by a current I' flowing in c', a circular coil $r=\epsilon$, $z=0$ where ϵ is small and

$$M=(0.4)\pi I'\pi\epsilon^2 \tag{41}$$

Hence the calculation of the flux $\Delta\psi$ for any coil C' can be reduced to the calculation of the mutual inductance of C', c'. This applies to any shape and orientation of C'. By a well-known reciprocity relation, this calculation may be reduced to the calculation of the field component H_z at the origin, due to a current flowing in C'.

If the initial field is itself caused by NI ampere turns in a coil of radius a' at a large distance d' from the center of the conductor sphere, then if a'/d' is small, the initial field is very nearly constant over the area of the sphere, and is given by

$$H_0=\frac{2NI\pi a'^2}{10d'^3} \tag{42}$$

Calculation of Losses

The losses caused by the eddy currents in the sphere may be found by integration of $|E_\varphi^2|/2\rho$ over the volume of the sphere, or by integration of the normal component of the Poynting vector over a surface enclosing the sphere. They may also be obtained from the induced emf in a "primary" coil which is used to generate the "initial" or "applied" field $H_0\mathbf{k}$. The last method is considered.

Assume that the initial field is generated by a "primary" current J in a very long solenoid S of radius $b>a$, with axis along the z-axis, and with N turns per unit z, where

$$(0.4)\pi NJ=H_0 \tag{43}$$

The additional linkage in the primary coil turn at z, due to the introduction of the sphere, may be found from equation 39, and is given by

$$\frac{\pi H_0a^3b^2(\xi+j\eta)}{(b^2+z^2)^{3/2}} \tag{44}$$

yielding for the whole solenoid

$$\Delta\psi_s=N\pi H_0a^3b^2(\xi+j\eta)\int_{-\infty}^{+\infty}\frac{dz}{(b^2+z^2)^{3/2}}$$

$$=N\pi H_02a^3(\xi+j\eta) \tag{45}$$

This leads to a net (added) induced emf

$$\Delta(\text{emf})=-j\omega10^{-8}\Delta\psi_S \tag{46}$$

From point of view of the primary circuit, the sphere is equivalent to a series impedance

$$Z=-\frac{\Delta(\text{emf})}{J}=j8\pi^2N^2\omega10^{-9}a^3(\xi+j\eta) \tag{47}$$

as if an added inductance

$$\Delta L=8\pi^2N^210^{-9}a^3\xi \text{ henry} \tag{48}$$

and a resistance

$$R=-8\pi^2N^210^{-9}\omega a^3\eta \text{ ohm} \tag{49}$$

were inserted in series with the solenoid coil. The net losses are

$$W=J^2R/2=25H_0^2a^3\omega(-\eta)10^{-7} \text{ watts} \tag{50}$$

Rotating Magnetic Field

Replacing the steady (pulsating) applied or initial field, equation 16 by a field rotating in the y-, z-plane.

The rotating field may be resolved into two pulsating stationary fields, 90 degrees out of phase, as described by

$$H=H_0(\mathbf{k}+j\mathbf{j}) \tag{51}$$

Applying superposition, the field for a stationary sphere is calculated, as a result of the added component $jH_0\mathbf{j}$, from equations 19, 20, 32–35 by proper permutation of the x, y, z co-ordinates and field components, and multiplication by j.

In particular it will be found that the losses are $2W$ where W is given by equation 50.

The torque exerted on the sphere may be obtained by integrating the moments of the interaction forces of **H** and the currents in proper solenoids, or from stresses $H^2/8\pi$ along and across tubes of force. Consideration of the latter shows that one may replace the sphere by proper dipoles and calculate the reactions on them. Noting equation 33 and replacing A by $a^3(\xi+j\eta)$ one has a dipole of moment

$$\mathbf{M}=-2AH_0(\mathbf{k}+j\mathbf{j})=2a^3H_0(\xi+j\eta)(\mathbf{k}+j\mathbf{j}) \tag{52}$$

Recalling the factor $\epsilon^{j\omega t}$ and taking real parts

$$M_x=0$$

$$M_z=2a^3H_0(-\xi\cos\omega t+\eta\sin\omega t)$$

$$M_y=2a^3H_0(\eta\cos\omega t+\xi\sin\omega t) \tag{53}$$

Equation 51 yields for the "real" initial field

$$H_x=0,\ H_y=-H_0\sin\omega t,\ H_z=H_0\cos\omega t \tag{54}$$

At the time $t=0$, the dipole of moment $2a^3H_0\sqrt{\xi^2+\eta^2}$ makes an angle $\tan^{-1}\eta/\xi$ with the negative z-axis. The field $H_z=H_0$ produces on it a torque

$$T=2H_0a^3\eta \text{ dynes} \tag{55}$$

about the positive x-axis. The same torque results for any t.

Sphere in Field of a Circular Coil

As a more complex example utilizing the solutions 10–15 for all n, the field described in the title is considered.

First consider in free space the "primary" field of a current I flowing in a circular coil. Let the coil be placed so that its axis is the z-axis and its center at $z=d$. Let (see Fig. 2)

$$\alpha=\tan^{-1}(b/d),\quad c=\sqrt{b^2+d^2} \tag{56}$$

be the semivertical angle of the cone subtended by the coil at the origin, and c its distance from the origin. This field of a circular coil in free space, when displacement currents are negligible, as well-known, can be expressed in a finite form involving elliptic integrals. The electric field $\mathbf{E}$ is axially symmetric and consists of a single component $\mathbf{E}=\mathbf{i}_\varphi E_\varphi$, $E_\varphi=E_\varphi(R,\theta)$ only. For the present application the following representation for $\mathbf{E}$ will be used

$$\mathbf{E}=\mathbf{i}_\varphi 2\pi j\omega 10^{-9} I \sin\alpha \sum_{n=1}^{\infty} P_n^{(1)}(\cos\alpha)\, P_n^{(1)}(\cos\theta) \begin{cases} \left(\dfrac{R}{c}\right)^{n} & \text{for } R<c \\ \left(\dfrac{c}{R}\right)^{n+1} & \text{for } R>c \end{cases} \tag{57}$$

where

$$P_n^{(1)}(\mu)=(1-\mu^2)^{1/2}dP_n(\mu)/d\mu \tag{58}$$

Indeed, the terms of equation 57 yield Cartesian field components

$$E_x=-\text{constant } P_n^{(1)}(\cos\theta)\sin\varphi \times (R^n, R^{-(n+1)})$$

$$E_y=\text{constant } P_n^{(1)}(\cos\theta)\cos\varphi \times (R^n, R^{-(n+1)}) \tag{59}$$

which agree with equations 10, and 11 for $m=1$.

The determination of the coefficients in equation 57 is carried out so as to obtain continuity in E at $R=c$ and a proper discontinuity in H_θ corresponding to the current distribution over $R=c$. For details foregoing refer to reference 3.

Now suppose that a conducting sphere of radius a is introduced in the foregoing field, with center at the origin. Inside the sphere $R<a$ the field E_φ may be obtained by superposition of terms

$$P_n^{(1)}(\cos\theta)S_n(\alpha R) \tag{60}$$

while outside this sphere the field E_φ can be represented as a sum of equation 57 and an additional field obtained from terms

$$P_n^{(1)}(\cos\theta)R^{-(n+1)} \tag{61}$$

the coefficients in these solutions are determined so that the net E_φ and H_θ are continuous at $R=a$. There results for E_φ for $R<a$

$$\frac{E_\varphi}{Ij\omega 2\pi 10^{-9}\sin\alpha} \sum_{n=1}^{\infty} \frac{(2n+1)P_n^{(1)}(\cos\alpha)P_n^{(1)}(\cos\theta)(a/c)^n S_n(\alpha R)/S_n(\alpha a)}{n(n+1)[n\mu+1+F_n(\alpha a)]} \tag{62}$$

where

$$F_n(u)=S'_n(u)/uS_n(u) \tag{63}$$

Outside the sphere, for $R>a$, the additional field must be added to the field of the coil in free space, is given by

$$\frac{\mathbf{E}_\varphi'}{Ij\omega 2\pi 10^{-9}\sin\alpha}=\sum_{n=1}^{\infty} \frac{[\mu(n+1)-1-F_n(\alpha a)]P_n^{(1)}(\cos\alpha)P_n^{(1)}(\cos\theta)(a/c)^n(a/R)^{n+1}}{n(n+1)[n\mu+1+F_n(\alpha a)]} \tag{64}$$

The additional induced emf in the circular coil due to the introduction of the sphere can be evaluated from equation 64 and is given by

$$\Delta(\text{emf})=2\pi c\sin\alpha E_\varphi=Ij\omega 10^{-9}2\pi c\sin^2\alpha \sum_{n=1}^{\infty} \frac{[\mu(n+1)-1-F_n(\alpha a)][P_n^{(1)}(\cos\alpha)]^2(a/c)^{2n+1}}{n(n+1)[n\mu+1+F_n(\alpha a)]} \tag{65}$$

Complex Permeability

It is known that by replacing μ by the complex permeability

$$\mu=\mu_0\epsilon^{-j\delta} \tag{66}$$

one may take into account magnetic hystersis, at least in an approximate manner. Indeed, on introducing the factor $\epsilon^{j\omega t}$ and taking real parts of

$$H=H_0\epsilon^{-j\omega t},\quad B=\mu_0H_0\epsilon^{(j\omega t-\delta)} \tag{67}$$

there results in the H,B-plane an elliptic hysteresis loop as shown in Fig. 3, with a "coercive force"

$$H_c=H_0\sin\delta \tag{68}$$

and a retentivity

$$B_r=H_0\mu_0\sin\delta \tag{69}$$

at the intersections with the axes. The area of the ellipse is

$$\int BdH=\pi\mu H_0^2\sin\delta \tag{70}$$

For different H_0, similar ellipses result.

While the ellipses furnish a poor approximation to the shapes of actual hysteresis loops, the advantages of the foregoing procedure lies in the retention of linearity of the field equations.

Substitution of 66 in equation 6 yields

$$\alpha^2=4\pi\sigma\mu_0\epsilon^{-j\delta}\omega 10^{-9}j \tag{71}$$

and it will be seen that α^2 is no longer pure imaginary but makes an angle $-\delta$ with the pure imaginary axis. As a result α is given not by equation 6 but by

$$\alpha=\sqrt{2}\,\epsilon^{j\left(\frac{\pi}{4}-\frac{\delta}{2}\right)}2\pi\sqrt{\mu_0 f\sigma 10^{-9}} \tag{72}$$

Except for this difference the calculation of the field disturbance caused by the sphere in alternating uniform magnetic field or in the field of a coil can be carried out as before. In particular, for a sphere in a uniform field the ratio $\Delta\psi/\psi_0=(\xi+j\eta)$ can be computed in the same manner. Both ξ and η will be affected by the introduction of δ.

As an example consider for the uniform field the case $\gamma=1$, $\mu=10(1-0.3j)$ so that $\delta=\tan^{-1}0.3=16.6992$ degrees. There results

$$\begin{aligned} \alpha a &=(1+j)\gamma\epsilon^{-j\delta/2} \\ &=\sqrt{2}\gamma\epsilon^{j\left(\frac{\pi}{4}-\frac{\delta}{2}\right)} \\ &=(1.1346+j0.8442)\gamma \end{aligned} \tag{73}$$

Letting $\gamma=1.00$, $\mu=10-3j$, $=10(1-0.3j)$, one obtains

$$\frac{\Delta\psi}{\psi_0}=1.5153-0.1964j$$

These values are included in Table I.

This is to be compared with the corresponding value $\Delta\psi/\psi_0=1.4923-0.0846j$ in absence of hysteresis, for $\mu=10$, $\gamma=1$.

Conducting Bodies of Nonspherical Shape

The calculation of the disturbance produced in an alternating magnetic field by the presence of conducting bodies of shape other than spherical is naturally more complicated, and only a few general remarks regarding it will be made.

Consider an oblong, axially symmetric body, say a prolate spheroid of semiaxes a, b, b. Resolve the initial field along the axis of symmetry and at right angles. If the long dimension is sufficiently long, one may as a first approximation, consider it infinitely long. The eddy currents for an infinitely long cylinder of radius a, both for a field parallel to its axis, and for a field, normal to its axis, can be solved. In the former case, for an applied field $H_0 \epsilon^{j\omega t}$, the net flux is given by

$$\frac{\Delta\psi}{\psi_0}=\frac{(\alpha a)I_0(\alpha a)}{2I_1(\alpha a)} \tag{74}$$

where $\psi_0=\pi a^2H_0$, and I_0, I_1 are the familiar Bessel functions of pure imaginary argument. For a field $H_0\epsilon^{j\omega t}$ normal to the cylinder axis the induced field is given by

$$\frac{E}{j\omega 10^{-8}H_0}=\cos\theta\,\frac{2aI_1(\alpha r)}{\dfrac{\alpha a}{\mu}I_0(\alpha a)+\left(1-\dfrac{1}{\mu}\right)I_1(\alpha a)}$$

where r, θ are polar co-ordinates in a plane normal to the axis and containing direction of the field H_0.

Actually the latter solution will apply for prolate ellipses even if a/b is not infinite; the former, however, has to be corrected, especially if μ is large. In the well-known solution of the (d-c) magnetization of an ellipsoid by means of a field parallel to its long axis, a high concentration of flux is obtained. This effect persists even with eddy currents, when the reluctance of the skin layer is appreciably smaller than that of the surrounding space.

A more exact treatment of the ellipsoid must be carried out, of course, by solving the field equations within the ellipsoid and outside it. This solution of these equations involves ellipsoidal harmonics and Mathieu functions, and falls outside the scope of the present paper.

References

1. NOTES ON RECENT RESEARCH IN ELECTRICITY AND MAGNETISM, J. J. Thompson. Oxford University Press, London, England, 1893, pp. 308–321, 362.

2. STATIC AND DYNAMIC ELECTRICITY (book), W. R. Smythe. McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1950, pp. 397.

3. *Ibid.*, pp. 273–75.

Effects of Nonlinearity on Propagation in Ionized Media

GEORGE I. COHN
MEMBER AIEE

THE PRESENCE of an ionized media in the neighborhood of radio communication systems can degrade performance by:

1. Insertion of attenuation and distortion in each channel.
2. Interference between channels caused by generation of harmonics and sum and difference frequencies.

This paper is concerned with the nonlinear effects of the media on the propagating signal. These nonlinear effects constitute the interference signals. Under certain conditions extremely small nonlinear effects can produce significant interference signals. For example, consider here a 1-kw transmitter at one frequency and a receiver capable of detecting 1 micromicrowatt tuned to one of the nonlinearly generated frequencies; a difference of 15 orders of magnitude results. Even if the nonlinearity of the media is so slight that only a millionth of a millionth of the transmitter power is converted to the frequency the receiver is tuned to a tremendous interference signal may result under appropriate conditions of proximity.

The development presented here is restricted to a plane polarized wave propagating through a media which in the absence of electromagnetic fields is homogeneous. An additional restriction is that there are no electric or magnetic fields present which are not directly caused by the propagating wave under consideration. In a practical situation these restrictions will probably not prevail with a consequent considerable change in the phenomena taking place.

Separation of Macroscopic and Microscopic Phenomena

The analysis is divided into two parts:

1. Determination of the macroscopic propagation phenomena in a nonlinear media with the constitutive parameters expressed in a generalized form suitable for a wide variety of nonlinear media.
2. Determination of the constitutive parameters for the ionized media of interest.

This method of subdivision not only makes the analysis more tractable, but also makes it more general in that any restrictions introduced in one part do not cause a corresponding loss of generality in the results of the other part.

There are various ways to interpret Maxwell's equations

$$\nabla\times\mathbf{E}=-\partial_t\mathbf{B} \tag{F1}$$

$$\nabla\times\mathbf{H}=+\mathbf{i}\partial_t\mathbf{D} \tag{A1}$$

$$\nabla\cdot\mathbf{D}=\rho \tag{1}$$

$$\nabla\cdot\mathbf{B}=0 \tag{2}$$

(Important equations which undergo many modifications but essentially conserve their identity throughout a substantial portion of their analysis will be designated by means of a letter followed by a number. The letter is chosen for its mnemonic value, such as: F for Faraday's law and A for Ampere's law. The number indicates the particular form. This equation labeling scheme facilitates keeping track of the development.) The two extreme points of view suitable respectively for the two subdivisions of the problem are discussed in the following.

MAXWELL'S MACROSCOPIC PHENOMENOLOGICAL VIEW

This view, fathered by Maxwell, presents a scientific description of the observable macroscopic (compared with the atomic constituents of the media) phenomena. What is happening to the fundamental charged particles composing the media in which the phenomena are taking place is not described in this view. This phenomenological theory accounts for the influence of the media, i.e., the fundamental particles composing the media, by means of factors called constitutive parameters, μ, ϵ and σ, which relate the fields in the manner

$$\mathbf{B}=\mu\mathbf{H} \tag{3}$$

$$\mathbf{D}=\epsilon\mathbf{E} \tag{4}$$

$$\mathbf{i}=\sigma\mathbf{E} \tag{5}$$

in isotropic media. It is convenient to

Paper 59-1145, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted April 3, 1959; made available for printing August 13, 1959.

GEORGE I. COHN is with the Illinois Institute of Technology, Chicago, Ill.

This research was supported in part by the U. S. Air Force, through Wright Air Development Center Communication and Navigation Laboratory via Cook Research Laboratories Division of Cook Electric Company.

express the constitutive parameters in the factored form

$$\mu=\mu_r\mu_v \tag{6}$$

$$\epsilon=\epsilon_r\epsilon_v \tag{7}$$

where μ_v and ϵ_v are the permeability and capacitivity of vacuum, whereas μ_r and ϵ_r are the relative permeability and capacitivity of the media with respect to vacuum. A consequence of the view which deserves explicit statement is that for the charge density appearing in equation 1

$$\rho=\text{free charge density only} \tag{8}$$

The effect on the field because of the equal numbers of oppositely charged bound particles constituting the media is accounted for by the relative capacitivity ϵ_r. The effects on the fields because of the motions of the charges constituting the media are taken into account in three ways:

1. The motion of the unbound charges is subsumed in the conduction current by means of the conductivity σ.
2. The translatory motion of the bound charges is subsumed in the electric displacement field by means of the relative capacitivity ϵ_r.
3. The rotational motion of the constituent charges is subsumed in the magnetic induction field by means of the relative permeability μ_r.

LORENTZ'S MICROSCOPIC EXHAUSTIVE VIEW

This view, generated by Lorentz, attempts to present an exhaustive description of all electrical phenomena in terms of the elementary positive and negative charges constituting the media. The microscopic description dispenses with the solid-material media concept and instead considers matter only as an aggregate of positive and negative charges (which actually constitute the media) in free space. The microscopic view does not see the forest (material media) for the trees (constituent positive and negative charges). Contrast this with the macroscopic view which does not see the trees (constituent positive and negative charges) for the forest (material media). In the microscopic view electromagnetic fields are still related by the same set of Maxwell's equations F1, A1, 1, 2, but since the view does not encompass material media but only charges in vacuum it is necessary to replace the auxiliary relations 3 and 4 by

$$\mathbf{B}=\mu_v\mathbf{H} \tag{9}$$

$$\mathbf{D}=\epsilon_v\mathbf{E} \tag{10}$$

and to replace the charge density appearing in equation 1 by

$$\rho=\text{total charge density} \tag{11}$$

and to replace the current density relation 5 by

$$\mathbf{i}=\sum_k n_k q_k \mathbf{u}_k \tag{12}$$

where n_k is the number of particles per unit volume having a charge q_k and a velocity $\mathbf{u}_k$.

Maxwell's equations relate the fields to the charges and currents. In the microscopic point of view the currents are in turn related to the motions of the charges. Since the charge motions are unknown it is necessary to add Lorentz's dynamical equation of motion which in turn relates these motions to the fields

$$\frac{dm_k\mathbf{u}_k}{dt}=q_k[\mathbf{E}+\mathbf{u}_k+\mathbf{B}] \tag{13}$$

The set of equations F1, A1, 2, 3, and 9–13 must be solved simultaneously for the dynamical behavior of the charges. From this information the phenomenological theory of the description can be obtained. It should be pointed out that Lorentz's microscopic theory does not meet with complete success because it does not encompass within its framework the quantum mechanical laws obeyed by the microscopic constituents. However, in many cases, particularly where the charge density is not too great as in electron beams and ionized gases the microscopic theory of Lorentz meets with considerable success. In the macroscopic phenomenological view, solution for the behavior of the charged particles is obviated because these quantities do not explicitly occur in the formulation.

Subsumption of the dynamical behavior of the constituent charged particles by means of the constitutive parameters is of immense practical importance in a wide variety of situations. This importance lies in the simplification resulting from the subdivision separating the macroscopic aspects from the microscopic aspects of the problem. In a wide variety of media the constitutive parameters can be adequately determined with relative ease by empirical means compared to a theoretical determination. For such media the theoretical determination of the constitutive parameters can be dispensed with and only the phenomenological theory is relevant.

For media such as ionized gases the direct measurement of the constitutive parameters is not such a straightforward task as for substances composed of neutral molecules because the space charges present and the boundary conditions imposed by the measuring equipment profoundly affect each other. To separate out what part of the measured results is due to the original constitutive parameters and what part is due to the disturbance introduced by the measuring equipment may well be a more difficult task than the theoretical determination of the constitutive parameters.

Propagation in Nonlinear Media

RESTRICTION TO PLANE POLARIZED WAVE

The macroscopic aspect of propagation of a plane polarized wave in an isotropic and uniform (in the absence of any fields) media in the absence of any externally applied fields other than the wave itself is investigated in this section. The reference axes are oriented so that: the x-axis is in the direction of the electric field. (Equations relating specific information about a particular important variable that occurs frequently throughout the development will be designated by a number followed by the letter representing the variable. For example equations 1E, 2E, etc. designate the equations providing the first, second, etc. pieces of information about the variable E. This equation labeling scheme facilitates keeping track of the development.)

$$\mathbf{E}=1_x E \tag{1E}$$

the y-axis is in the direction of the magnetic field

$$\mathbf{H}=1_y H \tag{1H}$$

and the z-axis is in the direction of propagation

$$\pi=1_z EH \tag{14}$$

Since there is only an x-component of electric field and a y-component of magnetic field Maxwell's induction equations reduce to

$$\partial_z E=-\partial_t \mu H \tag{A2}$$

$$-\partial_z H=\sigma E+\partial_t \epsilon E \tag{B2}$$

This investigation is restricted to the case for which nonlinearity enters only through the conductivity and in such a manner that the conductivity can be expressed in a rapidly converging power series in terms of the electric field intensity

$$\sigma=\sum_{k=0}^{\infty}\sigma_k E^k \tag{1σ}$$

Substituting this expansion for the conductivity into the induction laws and placing all terms linear in the field quantities on the left-hand sides and all nonlinear terms on the right-hand sides gives

$$\partial_z E+\mu\partial_t H=0 \tag{A3}$$

$$\partial_z H+\epsilon\partial_t E+\sigma_0 E=-\sum_{k=1}^{\infty}\sigma_k E^{k+1} \tag{B3}$$

RESOLUTION INTO PARTIAL FIELDS

This pair of nonlinear partial differential equations can be conveniently solved by the reversion method.[1,2] This method consists of reducing the original set of nonlinear equations to a sequence of sets of linear equations. This is accomplished by expanding the fields in terms of partial fields

$$E=\sum_{m=1}^{\infty}E_m \tag{2E}$$

$$H=\sum_{m=1}^{\infty}H_m \tag{2H}$$

in such a way that each set of partial fields E_m and H_m satisfy the following requirements

1. E_m and H_m depend only on partial fields of the same or lower order.
2. E_m and H_m satisfy linear partial differential equations.

Resolution of the fields into partial fields satisfying these requirements can be achieved with the aid of an ordering parameter[1] introduced in the following manner

$$E=\sum_{m=1}^{\infty}1^m E_m \tag{3E}$$

$$H=\sum_{m=1}^{\infty}1^m H_m \tag{3H}$$

Since

$$1^m=1 \tag{15}$$

Equations 3E and 3H are the same as equations 2E and 2H. By leaving the ordering parameters in the form introduced in equation 3E and 3H and temporarily treating the ordering parameter as a non-numeric the desired equivalent sets of linear equations can be obtained by substituting the partial field expansions into the nonlinear equations and equating coefficients of like powers of the ordering parameter. Substituting equations 3E and 3H into equations A3 and B3 gives

$$\sum_{m=1}^{\infty}1^m(\partial_z E_m+\mu\partial_t H_m)=0 \tag{A4}$$

$$\begin{aligned}\sum_{m=1}^{\infty}1^m(\partial_z H_m+\epsilon\partial_t E_m+\sigma_0 E_m)=\\ -\sigma_1[1^1E_1+1^2E_2+1^3E_3+\ldots.]^2\\ -\sigma_2[1^1E_1+1^2E_2+1^3E_3+\ldots.]^3\\ -\sigma_3[1^1E_1+1^2E_2+1^3E_3+\ldots.]^4\ldots\end{aligned} \tag{B4}$$

Equating coefficients of like powers of the ordering parameters gives

$$\partial_z E_m+\mu\partial_t H_m=0 \tag{A5}$$

$$\partial_z H_m+\epsilon\partial_t E_m+\sigma_0 E_m=-f_m(t,z) \tag{B5}$$

where

$$f_1(t,z)=0 \tag{16}$$

$$f_2(t,z)=\sigma_1 E_1^2 \tag{17}$$

$$f_3(t,z)=\sigma_2 E_1^3+2\sigma_1 E_1 E_2 \tag{18}$$

$$f_4(t,z)=\sigma_3 E_1^4+3\sigma_2 E_1^2 E_2+\sigma_1(2E_1E_3+E_2^2) \tag{19}$$

$$f_5(t,z)=\sigma_4 E_1^5+4\sigma_3 E_1^3 E_2+\sigma_2(3E_1^2E_3+3E_1E_2^2)+\sigma_1(2E_1E_4+2E_2E_3) \tag{20}$$

etc.

Note that E_1 and H_1 satisfy the same partial differential equations as the total fields satisfy in the case for which there is no nonlinearity, and hence, would constitute the solution to the original problem if the media were linear. All higher order partial fields satisfy the same pair of differential equations except that the equations are no longer homogeneous but are now driven by distributed excitation functions that depend on lower order partial fields.

RESOLUTION INTO FOURIER COMPONENTS

The analysis presented here is restricted to determination of steady-state time behavior. Since the partial differential equations satisfied by the partial fields are linear the partial fields will contain only those frequencies which are present in their driving functions. Examination of equations 16–20 etc. reveals that each driving function contains the frequencies of the first partial field plus the various harmonics, sums, and differences which can be obtained through products of the same order as the driving function. For example if the first partial fields consist of a single frequency the mth order driving function can have harmonic frequencies through the mth order only.

$$f_m=\Re\sum_{n=0}^{m}f_m^n\,e^{jn\omega t} \tag{21}$$

Consequently

$$E_m=\Re\sum_{n=0}^{m}E_m^n e^{jn\omega t} \tag{4E}$$

$$H_m=\Re\sum_{n=0}^{m}H_m^n e^{jn\omega t} \tag{4H}$$

where f_m^n, E_m^n, and H_m^n are respectively the complex amplitudes of the nth harmonics of the mth driving function, mth partial electric field, and mth partial magnetic field. This choice of complex Fourier representation separates each partial differential equation involving mth order partial fields into $m+1$ ordinary differential equations for the various frequency components

$$\frac{d}{dz}E_m^n+jn\omega\mu H_m^n=0\quad n=0,1,2,\ldots.,m \tag{A6}$$

$$\frac{d}{dz}H_m^n+(\sigma_0+jn\omega\epsilon)E_m^n=-f_m^n(z)\qquad n=0,1,2,\ldots.,m \tag{B6}$$

REDUCTION BY LAPLACE TRANSFORMS

These equations will be solved by means of the Laplace transform

$$\mathfrak{L}f(z)=\bar{f}(s)=\int_0^\infty e^{-sz}f(z)dz \tag{22}$$

which reduces them to the algebraic equations

$$s\bar{E}_m^n-E_{m0}^n+jn\omega\mu\bar{H}_m^n=0 \tag{A7}$$

$$s\bar{H}_m^n-H_{m0}^n+(\sigma_0+jn\omega\epsilon)\bar{E}_{m0}^n=-\bar{f}_m^n \tag{B7}$$

where

$$E_{m0}^n=E_m^n(0) \tag{5E}$$

$$H_{m0}^n=H_m^n(0) \tag{5H}$$

Solving equations A7 and B7 for the transforms of the nth harmonics in the mth partial fields gives

$$\bar{E}_m^n=\frac{sE_{m0}^n-jn\omega\mu(H_m^n-\bar{f}_m^n)}{s^2-\gamma_n^2} \tag{6E}$$

$$\bar{H}_m^n=\frac{s(H_{m0}^n-\bar{f}_m^n)-(\sigma_0+jn\omega\epsilon)E_{m0}^n}{s^2-\gamma_n^2} \tag{6H}$$

where

$$\gamma_n^2=jn\omega\mu(\sigma_0+jn\omega\epsilon) \tag{23}$$

has been used to abbreviate the equation writing. The symbol gamma has been chosen for this cluster of symbols because it plays the role of a propagation factor in exactly the same manner as for a linear media.

MAGNETIC FIELD AND CONSTANT COMPONENTS

Since the harmonic components of the partial magnetic field can be obtained directly from equation A6

$$H_m^n(z)=-\frac{1}{jn\omega\mu}\frac{d}{dz}E_m^n(z)n\geq 1 \tag{7H}$$

it is not necessary to invert equation 6H. The formula for the constant component must be developed separately since division by zero is not allowed. For this case equations A6 and B6 reduce to

$$\frac{dE_m^0}{dz}=0 \tag{8A}$$

$$\frac{dH_m^0}{dz}+\sigma_0E_m^0=-f_m^0(z) \tag{B8}$$

Solving gives

$$E_m^0(z)=E_{m0}^0 \tag{7E}$$

$$H_m^0(z)=H_{m0}^0-\sigma_0E_{m0}^0z-\int_0^z f_m^0(s)ds \tag{8H}$$

INVERSION OF LAPLACE TRANSFORM

Inversion of the transform of the *n*th harmonic of the *m*th partial electric field is facilitated by partial fraction expansion to

$$E_m^n=\frac{1}{2}\frac{E_n^m-Z_n(H_{m0}^n-\bar{J}_m^n)}{s-\gamma_n}+\frac{1}{2}\frac{E_{n0}^m+Z_n(H_{m0}^n-\bar{J}_m^n)}{s+\gamma_n} \tag{8E}$$

where

$$Z_n=\sqrt{\frac{jn\omega\mu}{\sigma_0+jn\omega\epsilon}}=\frac{jn\omega\mu}{\gamma_n} \tag{24}$$

Using

$$\mathfrak{L}^{-1}\frac{1}{s-\gamma_n}=e^{\gamma_nz} \tag{25}$$

and

$$\mathfrak{L}^{-1}\bar{f}(s)\bar{g}(s)=\int_0^z g(s-z)f(s)ds \tag{26}$$

to invert equation 8E gives

$$E_m^n(z)=\frac{E_{m0}^n-Z_nH_{m0}^n}{2}e^{\gamma_nz}+\frac{1}{2}Z_ne^{-\gamma_nz}\int_0^z e^{\gamma_ns}f_n^m(s)ds+\frac{E_{m0}^n+Z_nH_{m0}^n}{2}e^{-\gamma_nz}-\frac{1}{2}Z_ne^{-\gamma_nz}\int_0^z e^{-\gamma_ns}f_n^m(s)ds \tag{9E}$$

The excitation takes place over a plane boundary at *z* equal zero. The treatment is restricted to a single frequency excitation all of which is taken into account by the first partial field. consequently

$$E_{m0}^n=0 \qquad n\geq 2 \tag{10E}$$

$$H_{m0}^n=0 \qquad n\geq 2 \tag{9H}$$

As shown subsequently, nonlinearity produces negatively traveling *n*th harmonics which may make equations 10E and 9H untrue. With sufficiently small nonlinearity these are approximately true and the current development uses this approximation. Thus equation 9E simplifies to

$$E_m^n(z)=\frac{1}{2}Z_n\left[e^{-\gamma_nz}\int_0^z e^{\gamma_ns}f_m^n(s)ds-e^{\gamma_nz}\int_0^z e^{-\gamma_ns}f_m^n(s)ds\right] \tag{11E}$$

for $m\geq 2$

For the first order partial fields the distributed driving function is zero so equation 9E simplifies to

$$E_1^n(z)=\frac{E_{10}^n-Z_nH_{10}^n}{2}e^{\gamma_nz}+\frac{E_{10}^n+Z_nH_{10}^n}{2}e^{-\gamma_nz} \tag{12E}$$

FIRST ORDER PARTIAL FIELDS

Since there is no constant excitation applied

$$E_{10}^0=0 \tag{13E}$$

$$H_{10}^0=0 \tag{10H}$$

Consequently

$$E_1^0(z)=0 \tag{14E}$$

Since only a forward traveling wave of the fundamental frequency is being applied at the driving boundary of a medium which is homogeneous in the absence of fields it might be expected that no backward traveling wave would be present at this boundary so that

$$E_{10}^1-Z_1H_{10}^1=0 \tag{27}$$

Nonlinearity will cause a negatively traveling wave at the excitation boundary which makes equation 27 untrue. However, for sufficiently small nonlinearity equation 27 is an adequate approximation. Using equation 27 reduces equation 12E to

$$E_1^1(z)=E_{10}^1e^{-\gamma_1z} \tag{15E}$$

Using equations 7H and 24 gives

$$H_1^1(z)=\frac{E_{10}^1}{Z_1}e^{-\gamma_1z} \tag{11H}$$

Using

$$E_m^n(t,z)=\Re E_m^n(z)e^{jn\omega t} \tag{16E}$$

$$H_m^n(t,z)=\Re H_m^n(z)e^{jn\omega t} \tag{12H}$$

and

$$\gamma_n=\alpha_n+j\beta_n \tag{28}$$

gives for the fundamental components of the first partial fields

$$E_1^1(t,z)=E_{10}^1e^{-\alpha_1z}\cos(\omega t-\beta_1z) \tag{17E}$$

$$H_1^1(t,z)=\frac{E_{10}^1}{|Z_1|}e^{-\alpha_1z}\cos(\omega t-\beta_1z-\underline{/Z_1}) \tag{13H}$$

SECOND-ORDER PARTIAL FIELDS

Substituting equation 17E into 17 gives

$$f_2(t,z)=\sigma_1E_{10}^{1^2}e^{-2\alpha_1z}\cos^2(\omega t-\beta_1z) \tag{29}$$

Converting the nonlinear combination of trigonometric functions into a linear combination gives

$$f_2(t,z)=\frac{\sigma_1}{2}E_{10}^{1^2}e^{-2\alpha_1z}\times[1+\cos 2(\omega t-\beta_1z)] \tag{30}$$

Expressing this in the exponential form gives

$$f_2(t,z)=\Re\frac{\sigma_1}{2}E_{10}^{1^2}e^{-2\alpha_1z}[1+e^{j2(\omega t-\beta_1z)}] \tag{31}$$

Consequently

$$f_2^0(z)=\frac{\sigma_1}{2}E_{10}^{1^2}e^{-2\alpha_1z} \tag{32}$$

$$f_2^1(z)=0 \tag{33}$$

$$f_2^2(z)=\frac{\sigma_1}{2}E_{10}^{1^2}e^{-2\gamma_1z} \tag{34}$$

Substituting these into equation 11E gives

$$E_2^0(z)=0 \tag{18E}$$

$$E_2^1(z)=0 \tag{19E}$$

$$E_2^2(z)=\frac{\sigma_1}{4}Z_2E_{10}^{1^2}\left[e^{-\gamma_2z}\int_0^z e^{(\gamma_2-2\gamma_1)s}ds-e^{\gamma_2z}\int_0^z e^{-(\gamma_2+2\gamma_1)s}ds\right] \tag{20E}$$

Integrating and simplifying gives

$$E_2^2(z)=\frac{\sigma_1}{4}Z_2E_{10}^{1^2}\left[\frac{2\gamma_2e^{-2\gamma_1z}}{\gamma_2^2-4\gamma_1^2}-\frac{e^{-\gamma_2z}}{\gamma_2-2\gamma_1}-\frac{e^{\gamma_2z}}{\gamma_2+2\gamma_1}\right] \tag{21E}$$

Thus, the second harmonic is the only nonzero part of the second partial field. Using equation 16E gives

$$E_2^2(t,z)=\frac{\sigma_1}{4}E_{10}^{1^2}|Z_2|\left[\frac{2|\gamma_2|}{|\gamma_2^2-4\gamma_1^2|}e^{-2\alpha_1z}\cos(2\omega t-2\beta_1z+\underline{/\gamma_2}-\underline{/\gamma_2^2-4\gamma_1^2}+\underline{/Z_2})-\frac{e^{-\alpha_2z}}{|\gamma_2-2\gamma_1|}\cos(2\omega t-\beta_2z-\underline{/\gamma_2-2\gamma_1}+\underline{/Z_2})-\frac{-e^{\alpha_2z}}{|\gamma_2+2\gamma_1|}\cos(2\omega t+\beta_2z-\underline{/\gamma_2+2\gamma_1}+\underline{/Z_2})\right] \tag{22E}$$

For small nonlinearities of the type discussed here equation 22E may represent the principal effect of the nonlinearity on the electric field. This contribution by the nonlinear media consists of three different traveling waves. The first wave has a phase velocity

$$v_{p2,1}=\frac{2\omega}{2\beta_1}=\frac{\omega}{\beta_1}=v_{p1} \tag{35}$$

which is the same as that of the fundamental. The second wave has a phase velocity

$$v_{p2,2}=\frac{2\omega}{\beta_2} \tag{36}$$

which is different from that of the first and the third wave has a phase velocity

$$v_{p2,3}=\frac{2\omega}{-\beta_2}=-v_{p2,2} \tag{37}$$

and hence is a negatively traveling wave. The first wave attenuates at twice the rate of the fundamental and hence soon becomes negligible compared with it. In order to make meaningful statements about the relative importance of the waves it is necessary to determine their relative amplitudes, attenuation factors and phase velocities. One very interesting feature which can be emphasized at this point is that the negatively traveling wave is increasing in the positive z-direction. For a media of indefinitely great extent this formula gives an ever increasing value for this wave as z increases. This is physically impossible. It is necessary to keep in mind that the higher order partial fields are negligible only so long as E_2 is not too large, consequently equation 22E is not valid for excessively large values of z.

In the principal applications of interest here the conduction current is generally small compared with the displacement current

$$\sigma_0 \ll \omega\epsilon \tag{2\sigma}$$

Consequently it is convenient to express the propagation factor in the form

$$\gamma_n=jn\omega\sqrt{\mu\epsilon}\sqrt{1+\frac{\sigma_0}{jn\omega\epsilon}} \tag{38}$$

which is suitable for expansion giving

$$\gamma_n=jn\omega\sqrt{\mu\epsilon}\left[1+\frac{1}{2}\frac{\sigma_0}{jn\omega\epsilon}-\frac{1}{8}\left(\frac{\sigma_0}{jn\omega\epsilon}\right)^2+\frac{1}{16}\left(\frac{\sigma_0}{jn\omega\epsilon}\right)^3+\dots\right] \tag{39}$$

Separating into real and imaginary parts gives

$$\alpha_n=\frac{\sigma_0}{2}\sqrt{\frac{\mu}{\epsilon}}\left[1-\frac{1}{2}\left(\frac{\sigma_0}{2n\omega\epsilon}\right)^2+\dots\right] \tag{40}$$

$$\beta_n=\frac{n\omega}{c}\left[1+\frac{1}{2}\left(\frac{\sigma_0}{2n\omega\epsilon}\right)^2+\dots\right] \tag{41}$$

Since the second term in each expansion is small compared to the first it will be neglected where sums are involved. However, in expressions involving differences it is necessary to consider the first two terms since the first terms alone cancel. Using this degree of approximation gives

$$\alpha_2 \simeq \alpha_1 \tag{42}$$

$$\beta_2 \simeq 2\beta_1\left[1-\frac{3}{8}\frac{\alpha_l^2}{\beta_l^2}\right] \tag{43}$$

$$\frac{2|\gamma_2|}{|\gamma_2^2-4\gamma_1^2|} \simeq \frac{1}{\alpha_l} \tag{44}$$

$$|\gamma_2-2\gamma_1| \simeq \alpha_l \tag{45}$$

$$|\gamma_2+2\gamma_1| \simeq 4\beta_l \tag{46}$$

$$|Z_2| \simeq \sqrt{\frac{\mu}{\epsilon}} \tag{47}$$

$$\underline{/\gamma_2} \simeq \frac{\pi}{2} \tag{48}$$

$$\underline{/\gamma_2^2-4\gamma_1^2} \simeq -\frac{\pi}{2} \tag{49}$$

$$\underline{/\gamma_2-2\gamma_1} \simeq 0 \tag{50}$$

$$\underline{/\gamma_2+2\gamma_1} \simeq \frac{\pi}{2} \tag{51}$$

$$\underline{/Z_2} \simeq 0 \tag{52}$$

where

$$\alpha_l=\frac{\sigma_0}{2}\sqrt{\frac{\mu}{\epsilon}} \tag{53}$$

$$\beta_l=\omega\sqrt{\mu\epsilon} \tag{54}$$

are the attenuation and phase factors for the corresponding linear media.

Using these approximations simplify equation 22E to

$$E_2^2(t,z)=\frac{\sigma_1}{2\sigma_0}E_{10}^{1^2}\left[e^{-2\alpha_1 z}\cos(2\omega t-\alpha\beta_1 z)-e^{-\alpha_1 z}\cos(2\omega t-\beta_2 z)-\frac{\alpha_l}{\beta_l}e^{\alpha_1 z}\sin(2\omega t+\beta_2 z)\right] \tag{23E}$$

If α_1 is sufficiently small it might be expected that the beat phenomena between the two forward wave components would be important. The beat envelope wavelength

$$\lambda_e=\frac{2\pi}{\beta_1-\frac{\beta_2}{2}} \tag{55}$$

to the degree of approximation used is

$$\lambda_e=\frac{2\pi}{\beta_1}\frac{8}{3}\frac{\beta_l^2}{\alpha_l^2} \tag{56}$$

or in normalized form

$$\frac{\lambda_e}{\lambda}=\frac{8}{3}\frac{\beta_l^2}{\alpha_l^2} \tag{57}$$

Using the attenuation length

$$L_\alpha=\frac{1}{\alpha_l} \tag{58}$$

and the wavelength

$$\lambda=\frac{2\pi}{\beta_l} \tag{59}$$

gives

$$\frac{\lambda_e}{\lambda}=\frac{2}{3}\left(4\pi\frac{L_\alpha}{\lambda}\right)^2 \tag{60}$$

Consequently the beat wavelength is much too large compared with the attenuation length for the beat phenomena to be significant. The beat wavelength will be greater than the attenuation length only when the attenuation length is small compared with a wavelength in which case the wave will be damped out too rapidly for significant beats to occur.

Conductivity of an Ionized Media

Rather than express the current density in terms of each individual charged particle it is more convenient to lump all the charged particles of the same type and velocity into a single specie, and express the current density as the sum of the current densities due to each specie

$$\mathbf{i}=\sum_{s=1}^{S}\mathbf{i}^s \tag{61}$$

The current density due to each specie is

$$\mathbf{i}^s=n^s q^s \mathbf{u}^s \tag{62}$$

where

q^s = charge per ion of the sth specie
$\mathbf{u}^s$ = drift velocity of the ions of the sth specie
n^s = number of sth specie ions per unit volume

A more exact treatment than given in the following requires that the current density sum be expressed in terms of classes

$$\mathbf{i}=\sum_{c=1}^{C}\mathbf{i}^c \tag{63}$$

where the current density for a single class is due to all the ions of that type and is given by the integral of the velocity distribution function

$$\mathbf{i}^c=q^c\int\frac{d\mathbf{u}^c}{dn}\,dn \tag{64}$$

This more exact statistical mechanics treatment is beyond the scope of this paper. The objective of this section is to pursue a sufficiently simple heuristic approach so that the delineation of the nature of the inherent nonlinearities and a determination of their order of magnitude will be easily obtained.

The drift velocity of the sth specie is

$$\mathbf{u}^s=\frac{1}{2}\mathbf{a}^s\tau^s \tag{65}$$

where

$\mathbf{a}^s$ = acceleration of the sth specie

τ^s = mean free time between collisions of the *s*th specie

In the absence of magnetic fields the acceleration is given by

$$\mathbf{a}^s=\frac{q^s}{m^s}\mathbf{E} \qquad (66)$$

The mean free time in terms of the collision frequency is

$$\tau^s=\frac{1}{\nu^s} \qquad (67)$$

The drift velocity in terms of the electric field and the collision frequency is therefore

$$\mathbf{u}^s=\frac{q^s}{2m^s\nu^s}\mathbf{E} \qquad (68)$$

Using this expression for the drift velocity in equation 62 gives

$$\mathbf{i}^s=\frac{n^s(q^s)^2}{2m^s\nu^s}\mathbf{E} \qquad (69)$$

Since the current density is related to the electric field by the conductivity

$$\mathbf{i}^s=\sigma^s\mathbf{E} \qquad (70)$$

the expression for the conductivity due to the *s*th specie is

$$\sigma^s=\frac{n^s(q^s)^2}{2m^s\nu^s} \qquad (3\sigma)$$

Using equations 5 and 61 gives the conductivity addition law

$$\sigma=\sum_{s=1}^{S}\sigma^s \qquad (4\sigma)$$

The media can be nonlinear via the conductivity if any of the four parameters n^s, m^s, q^s, and ν^s vary with the field intensity. In the treatment as formulated in the foregoing the charge per ion of the *s*th specie is constant by definition. However, if the formulation is based on defining the specie in terms of the mass then m^s would be constant by definition and q^s could be a variable dependent on the electric field intensity.

COLLISION FREQUENCY DEPENDENCE ON ELECTRIC FIELD INTENSITY

The velocity of an ion can be resolved into thermal dependent and electric-field dependent parts

$$\mathbf{C}^s=\mathbf{C}_T^s+\mathbf{C}_E^s \qquad (71)$$

where

$\mathbf{C}_T^s$ = ion velocity component due to the thermal motion

$\mathbf{C}_E^s$ = ion velocity component due to the accelerating electric field

For the case of isotropic scattering

$$\mathbf{C}_E^s=\frac{q^s}{m^s}\mathbf{E}(t-t_c) \qquad (72)$$

where

t_c = time of the preceding collision

Taking the scalar product of $\mathbf{C}^s$ with itself gives

$$C^{s2}=C_T^{s2}+2\mathbf{C}_T^s\cdot\mathbf{C}_E^s+C_E^{s2} \qquad (73)$$

Since $\mathbf{C}_T^s$ is randomly oriented and $\mathbf{C}_T^s$ is in the same direction for all ions in a small but not too small region of space

$$\overline{\mathbf{C}_T^s\cdot\mathbf{C}_E^s}=0 \qquad (74)$$

Consequently, taking the average value of both sides of equation 73 gives

$$\overline{C^{s2}}=\overline{C_T^{s2}}+\overline{C_E^{s2}} \qquad (75)$$

Neglecting the dependence of the mean free path λ^s on the electric field intensity and dividing both sides by the square of the mean free path gives

$$\frac{\overline{C^{s2}}}{\lambda^{s2}}=\frac{\overline{C_T^{s2}}}{\lambda^{s2}}+\frac{\overline{C_E^{s2}}}{\lambda^{s2}} \qquad (76)$$

Approximating the squares of the collision frequencies by

$$\nu^{s2}=\frac{\overline{C^{s2}}}{\lambda^{s2}} \qquad (77)$$

and

$$\lambda_T^{s2}=\frac{\overline{C_T^{s2}}}{\nu^{s2}} \qquad (78)$$

gives

$$\nu^{s2}=\nu_T^{s2}+\frac{\overline{C_E^{s2}}}{\lambda^{s2}} \qquad (79)$$

When the second term on the right-hand side of equation 79 is very small compared with the first term, as is usually the case, the residual approximation error in equation 79 is much smaller than the approximation errors in equations 77 and 78. Substituting equation 72 into the formula for the mean square

$$\overline{C_E^{s2}}=\frac{1}{\tau}\int_{t_c}^{t_c+\tau^s}C_E^{s2}\,dt \qquad (80)$$

Integrating and expressing in terms of collision frequency gives

$$\overline{C_E^{s2}}=\frac{1}{3}\left(\frac{q^sE}{m^s\nu^s}\right)^2 \qquad (81)$$

Substituting this into equation 79 gives

$$\nu^{s2}=\nu_T^{s2}+\frac{1}{3}\left(\frac{q^sE}{m^s\lambda^s\nu^s}\right)^2 \qquad (82)$$

Expressing this equation as a conventional quadratic in terms of the collision frequency squared gives

$$\nu^{s4}-\nu_T^{s2}\nu^{s2}-\frac{1}{3}\left(\frac{q^sE}{m^s\lambda^s}\right)^2=0 \qquad (83)$$

Solving for the square of the collision frequency gives

$$\nu^{s2}=\frac{1}{2}\nu_T^{s2}\pm\sqrt{\frac{1}{4}\nu_T^{s4}+\frac{1}{3}\left(\frac{q^sE}{m^s\lambda^s}\right)^2} \qquad (84)$$

The negative sign is extraneous since the collision frequency and its square cannot be negative. Taking the root gives

$$\nu^s=\frac{1}{\sqrt{2}}\nu_T^s\sqrt{1+\sqrt{1+\frac{1}{3}\left(\frac{2q^sE}{m^s\lambda^s\nu_T^{s2}}\right)^2}} \qquad (85)$$

The average energy gained by the ion from the electric field between collisions is

$$W_E^s=q^sE\lambda^s \qquad (86)$$

The average thermal kinetic energy is

$$W_T^s=\frac{1}{2}m^s\lambda^{s2}\nu_T^{s2} \qquad (87)$$

Consequently equation 85 can be conveniently expressed in the form

$$\nu^s=\frac{1}{\sqrt{2}}\nu_T^s\sqrt{1+\sqrt{1+\frac{1}{3}\left(\frac{W_E^s}{W_T^s}\right)^2}} \qquad (88)$$

If the average energy gained from the electric field is small compared with the average thermal kinetic energy

$$\left(\frac{W_E^s}{W_T^s}\right)^2<<1 \qquad (89)$$

An approximation for equation 88 is

$$\nu^s=\nu_T^s\left[1+\frac{1}{24}\left(\frac{W_E^s}{W_T^s}\right)^2\right] \qquad (90)$$

or in terms of the electric field

$$\nu^s=\nu_T^s\left[1+\frac{1}{6}\left(\frac{q^sE}{m^s\lambda^s\nu_T^{s2}}\right)^2\right] \qquad (91)$$

Utilizing this result in equation 3σ and simplifying to the same degree of approximation gives

$$\sigma^s=\frac{n^sq^{s2}}{2m^s\nu_T^s}\left[1-\frac{1}{6}\left(\frac{q^sE}{m^s\lambda^s\nu_T^{s2}}\right)^2\right] \qquad (5\sigma)$$

Consequently

$$\sigma_0^s=\frac{n^sq^{s2}}{2m^s\nu_T^s} \qquad (6\sigma)$$

$$\sigma_1^s=0 \qquad (7\sigma)$$

$$\sigma_2^s=-\frac{\sigma_0^s}{6}\left(\frac{q^s}{m^s\lambda^s\nu_T^{s2}}\right)^2 \qquad (8\sigma)$$

For electrons

$$\frac{q_e}{m^e}=1.76\times10^{11}\ \text{coulomb/kg/(kilogram)} \qquad (92)$$

in nitrogen at a pressure of

$$p=76\ \text{millimeters mercury} \qquad (93)$$

at a temperature of

$$T = -54 \text{ degrees centigrade} \tag{94}$$

the mean free path and collision frequency are approximately

$$\lambda^{e} = 3.84 \times 10^{-6} \text{ m (meter)} \tag{95}$$

and

$$\nu^{e} = 2.4 \times 10^{10}/\text{second} \tag{96}$$

Consequently

$$\frac{\sigma_2^{e}}{\sigma_0^{e}} = -1.06 \times 10^{-11} \text{ m}^2/\text{volt}^2 \tag{9σ}$$

For the examples such as given in the introduction this could lead to undesirable interference under a variety of conditions.

For singly ionized nitrogen atoms

$$\frac{q^{N\pm}}{m^{N\pm}} = \pm 6.85 \times 10^{6} \text{ coulomb/kg} \tag{97}$$

In nitrogen at the foregoing pressure and temperature

$$\lambda^{N\pm} = 7 \times 10^{-7} \text{ m} \tag{98}$$

and

$$\nu^{N\pm} = 8 \times 10^{8}/\text{second} \tag{99}$$

Consequently

$$\frac{\sigma_2^{N\pm}}{\sigma_0^{N\pm}} = -4 \times 10^{-13} \text{ m}^2/\text{volt}^2 \tag{10σ}$$

The contribution of these heavy ions to the nonlinearity can be relatively important even though this ratio is so much smaller than that for the electrons because of the much greater densities of the heavy ions which may be present under various circumstances.

Ion Mass Dependence on Electric Field Intensity

The principal phenomena of interest here occur at electric field intensities sufficiently low that relativistic mass variation and mass variation due to the creation of new ion pairs can be neglected by comparison. The principal phenomena of importance here are caused by labile-cluster ions. Addition of neutral molecules to ions form heavier ions which are called labile-cluster ions. This clustering process can form very heavy ions. At each collision some of the neutral molecules or atoms may be knocked out of the labile cluster thereby changing its mass. The greater the electric field intensity the more energetic the collisions are and the greater the number of neutral particles knocked out of the labile cluster per collision on the average. Consequently, the average value of the labile-cluster ion mass depends on the electric field intensity, the larger the electric field intensity the smaller the average mass.

As an example consider nitrogen at ordinary temperatures with a small percentage of the molecules ionized. For low accelerating electric field intensities the most abundant positive ion is N_4^+ and the next most abundant is N_2^+. In addition N_1^+ as well as N_6^+ and heavier ions are present but in less abundance. As the electric field intensity is increased the collisions become more violent and the relative abundance of the lighter ions increases at the expense of the relative abundance of the heavier ions. The conductivity due to the singly charged positive nitrogen ions, from equation 3σ, is

$$\sigma^{N^+} = \frac{n^{N^+} e^2}{2m^{N^+} \nu^{N^+}} \tag{11σ}$$

The average mass of the positive nitrogen ions is given by

$$m^{N^+} = \sum_{g=1}^{G} p^{N_g^+} m^{N_g^+} \tag{100}$$

where

$m^{N_g^+}$ = mass of the positive nitrogen ion containing g nitrogen atoms

$p^{N_g^+}$ = percentage of the time that, or the probability that, a positive nitrogen ion contains g nitrogen atoms

This probability depends on the electric field intensity; the power series expansion is

$$p^{N_g^+}(E) = \sum_{k=0}^{\infty} p_k^{N_g^+} E^k \tag{101}$$

Substituting this into equation 100 gives

$$m^{N^+} = \sum_{k=0}^{\infty} m_k^{N^+} E^k \tag{102}$$

where

$$m_k^{N^+} = \sum_{g=1}^{G} p_k^{N_g^+} m^{N_g^+} \tag{103}$$

Substituting equation 102 into 11σ gives

$$\sigma^{N^+} = \sum_{k=0}^{\infty} \sigma_k^{N^+} E^k \tag{12σ}$$

where

$$\sigma_0^{N^+} = \frac{n^{N^+} e^2}{2m_0^{N^+} \nu^{N^+}} \tag{13σ}$$

$$\sigma_1^{N^+} = -\sigma_0^{N^+} \frac{m_1^{N^+}}{m_0^{N^+}} \tag{14σ}$$

$$\sigma_2^{N^+} = \sigma_0^{N^+} \left[-\frac{m_2^{N^+}}{m_0^{N^+}} + \left(\frac{m_1^{N^+}}{m_0^{N^+}} \right)^2 \right] \tag{15σ}$$

$$\vdots$$

etc.

In this development the effect of the variable labile-cluster ion mass on the collision frequency has been neglected.

A good approximation to the average mass, equation 100, may be given by the contributions from the two most abundant nitrogen ions

$$m^{N^+} = p^{N_2^+} m^{N_2^+} + p^{N_4^+} m^{N_4^+} \tag{104}$$

since

$$p^{N_2^+} >> p^{N_g^+} \qquad g \neq 2,4 \tag{105}$$

and

$$p^{N_4^+} >> p^{N_g^+} \qquad g \neq 2,4 \tag{106}$$

However, this may be inadequate for the determination of the nonlinearity because of the probability that

$$p_k^{N_g^+} >> p_k^{N_2^+} \qquad \text{k} > 1 \text{ and } g \neq 2,4 \tag{107}$$

and

$$p_k^{N_6^+} >> p_k^{N_4^+} \qquad k > 1 \text{ and } g \neq 2,4 \tag{108}$$

That is, relatively small numbers of massive labile-cluster ions may give much greater contributions to the nonlinearity than the relatively large numbers of lighter ions because of the much greater sensitivity of their mass to the electric field intensity.

Ion Density Dependence on Electric Field Intensity

The phenomena described in the previous section can be treated in a more appropriate fashion by redefining the specie so that specification of the ion mass is necessary for specification of the specie. In this way the mass of the ion is no longer a variable and the ion density becomes a variable instead. For the positive nitrogen ions discussed in the preceding section

$$\sigma^{N^+} = \sum_{g=1}^{G} \sigma^{N_g^+} \tag{16σ}$$

where

$$\sigma^{N_g^+} = \frac{n^{N_g^+} e^2}{2m^{N_g^+} v^{N_g^+}} \tag{17σ}$$

The density of each ion specie depends on the electric field intensity, the power series expansion is

$$n^{N_g^+} = \sum_{k=0}^{\infty} n_k^{N_g^+} E^k \tag{109}$$

Substituting equations 109 into 17σ into 16σ again gives 12σ but now

$$\sigma_0^{N^+} = \frac{e^2}{2}\sum_{g=1}^{G}\frac{n_0^{N_g^+}}{m^{N_g^+}\nu^{N_g^+}} \qquad (18\sigma)$$

$$\sigma_1^{N^+} = \frac{e^2}{2}\sum_{g=1}^{G}\frac{n_1^{N_g^+}}{m^{N_g^+}\nu^{N_g^+}} \qquad (19\sigma)$$

$$\sigma_2^{N^+} = \frac{e^2}{2}\sum_{g=1}^{G}\frac{n_2^{N_g^+}}{m^{N_g^+}\nu^{N_g^+}} \qquad (20\sigma)$$

$$\vdots$$

etc.

A good approximation to $\sigma_0^{N^+}$ may be obtained by taking into account only the two most abundant specie

$$\sigma_0^{N^+} = \frac{e^2}{2}\left[\frac{n_0^{N_2^+}}{m^{N_2^+}\nu^{N_2^+}} + \frac{n_0^{N_4^+}}{m^{N_4^+}\nu^{N_4^+}}\right] \qquad (21\sigma)$$

The corresponding approximation is likely to be inadequate for determination of the nonlinearity because even relatively small numbers of the massive labile-cluster ions may cause substantial nonlinearity because of the weak bonds holding the added neutral molecules.

Conclusions

Steady-state propagation of a plane polarized wave in a homogeneous (in the absence of fields) nonlinear conducting media containing no electric or magnetic fields except those due directly to the wave has been investigated. Detailed results have been obtained for the case of single frequency excitation through quadratic terms in the electric field intensity. The principal nonlinear effects noted for single frequency excitation are the generation of three double frequency waves; two propagating in the forward direction and one propagating in the backward direction. One forward propagating wave travels with the same phase velocity as the fundamental but attenuates at twice the rate. The other forward propagating wave travels with a phase velocity slightly greater than that of the fundamental and attenuates at a slightly lower rate. The backward propagating wave has the same magnitude of phase velocity as the second forward propagating wave and increases exponentially with distance (for moderate distances) at the same rate as the second forward wave decreases with distance.

For an ionized gas the nonlinearity is due to the dependence of the collision frequency, the ion masses and the ion densities on the electric field intensity. A first-order approximation for the collision frequency showed that although the dependence on the electric field is very small it may still be significant in terms of producing an objectionable amount of nonlinearity. The variable ion densities or masses of labile-cluster ions are expected to cause much greater nonlinearity than that caused by the collision frequency dependence on the electric field. Continued investigation is required to obtain an estimation of this effect.

References

1. SOLUTION OF NONLINEAR DIFFERENTIAL EQUATIONS BY THE REVERSION METHOD, George I. Cohn, Bernard Saltzberg. *Journal of Applied Physics*, New York, N. Y., vol. 24, no. 2, Feb. 1953, pp. 180–86.

2. THE FIRST THIRTEEN TERMS IN THE SOLUTION OF NONLINEAR DIFFERENTIAL EQUATIONS BY THE REVERSION METHOD, George I. Cohn, Bernard Saltzberg. *Ibid.*, vol. 25, no. 2, Feb. 1954, pp. 252–54.

Transmission Aspects of an Electronic Switchboard Employing Time-Division Multiplexing

J. C. PERKINS, JR.
ASSOCIATE MEMBER AIEE

THE AN/TTC-12 switchboard is a newly designed electronic switchboard for military application employing solid-state switching techniques. This paper describes the transmission characteristics and performance of the system with respect to the multiplex methods employed. The systems logic and design philosophy of the switchboard are thoroughly described in reference 1.

Multiplex Method

Undoubtedly the major problem which confronts the design of any time-division multiplex (TDM) system is that of sampling loss. In a conventional system, each multiplexed channel will experience an insertion loss which is inversely proportional to the duty cycle τ/T, where τ is the duration of the individual samples or time-slot pulse width, and T is the time between samples or frame period. To achieve a large number of channels on a common transmission medium it is necessary that τ be made small, and T as large as possible consistent with the Nyquist criterion that the repetition frequency $1/T$ be at least twice as great as the highest signal frequency of the channel. Obviously, these requirements lead to high multiplex losses which in turn introduce the attendant problems of high-gain amplifier stability and signal-to-noise problems.

In the recent publications of Bennett[2] and Desoer,[3] a method for resonant transferring of stored signal energy through periodically operating switches with little or no loss is described. A simplified description of this principle, known as "resonant transfer," is illustrated in the circuit of Fig. 1(A).

Assume that the initial conditions of the circuit are such that switch S is open and a charge, V_1, exists on capacitor C_1. With closure of the switch, an oscillatory current i, whose frequency is determined by the LC product of the series resonant circuit, will flow. If the time the switch remains closed, τ, is selected to be half the natural period of oscillation, the conditions shown in Fig. 1(B) will exist. The energy in C_1 will have decreased to zero, the current having completed one-half cycle, and the total charge now ap-

---

Paper 59-210, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 24, 1958; made available for printing March 31, 1959.

J. C. PERKINS, JR. is with the Stromberg-Carlson Company, Rochester, N. Y.

The author acknowledges the many contributions made by the engineers of the Telecommunication Department, Stromberg-Carlson Company, in particular S. B. Weiner, W. H. Black, and D. A. Perreault. Acknowledgments are also extended to the members of the Switching Branch, Data Processing Division, Communications Department, U. S. Armed Services Research and Development Laboratories, for their assistance during the development phases of this project.

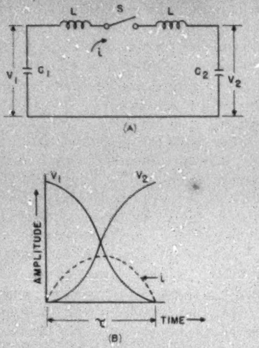


Fig. 1. Charge transfer of a resonant circuit

A—Resonant transfer circuit
B—Voltage-current relationships
Transfer conditions, resonant frequency: $\omega_r = 1/\sqrt{LC} = \pi/\tau$ or $\tau = \pi\sqrt{LC}$, where $C_1 = C_2 = C$

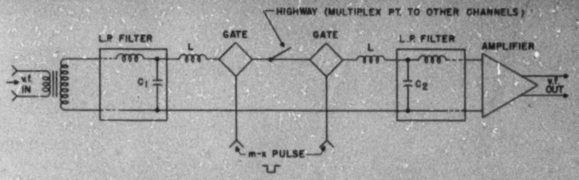


Fig. 2. Basic multiplex path of PAM system utilizing resonant transfer circuit

pears as V_2 across capacitor C_2. Thus an energy transfer has been completed between capacitors with no loss, assuming, of course, ideal components and zero switch resistance.

In application to a TDM system, the switch S is replaced by periodic sampling gates and the capacitors become end elements of the conventional low-pass filters used in a pulse-amplitude modulated (PAM) system to recover the intelligence. With the resonant transfer scheme as a base, the fundamental building blocks of a multiplex path become, as shown in Fig. 2, transmitting filter and gate, common transmission medium (highway), receiving gate and filter, and low-gain amplifier to overcome minor losses in the transfer process, filters, transformers, etc.

System Requirements

The selection of the sampling time τ frame time T, and the subsequent design criteria for the individual blocks in the multiplex path was dictated largely by the transmission specifications of the switchboard and the resonant-transfer method of time division.

In addition to having adequate bandwidth for voice transmission, each channel had to be capable of handling data information, which resulted in a phase-delay distortion specification of 10 μsec (microseconds) maximum per board. This stringent requirement was necessary since a connection could be extended through ten boards in tandem. A frame repetition frequency of 12.5 kc was selected for a theoretical maximum bandwidth of 6.25 kc of which approximately 5 kc would be attainable practically. This usable bandwidth was necessary to provide maximum phase linearity of the filters in their pass band to meet the delay requirement.

The allotted time for individual channels of 2.5 μsec in each frame period consisting of 80 μsec, provides a total of 32 channels per multiplex highway. As is the usual case in engineering design, this selection was a compromise of conflicting factors. A binary number of channels was desirable to simplify the switching circuitry. In addition, the number of channels per highway should be as large as possible to reduce the total equipment requirements. The transmission specifications, however, required an insertion loss stability of ±0.5 db (decibels) between any two connections through the board and a crosstalk ratio of greater than 50 db.

The power loss of resonant-transfer multiplexing is dependent on the effective pulse width of the combined transmit and receive gates; therefore, a wide sampling pulse is desirable to reduce level variations between channels as a result of differences in delay and the rise and fall times of the transistor circuitry. The channel time slot of 2.5 μsec was equally divided between sampling and guard time. A long guard time is desirable to allow signal energy stored in the distributed capacity of the highway to decay before the adjacent channel sample occurs; however, a clamping circuit was devised, as will be subsequently discussed, which insures a rapid dissipation of this crosstalk energy within the allotted time.

The output of any line circuit in the switchboard was required to contain no carrier leak or other frequency component produced by the modulation processes at a level greater than 30 db below the reference of 0 dbm (db referred to 1 milliwatt). Double polarity sampling, which balances out the repetition frequency and its harmonics, was chosen as

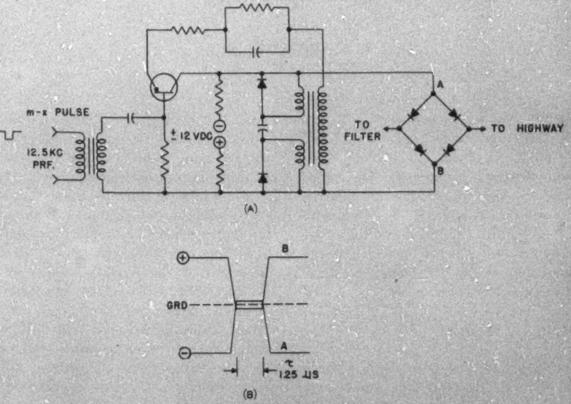


Fig. 3. Diode bridge gate operated from transistor blocking oscillator

A—Blocking oscillator gate
B—Voltage across diode bridge

Vol
Pt.
195

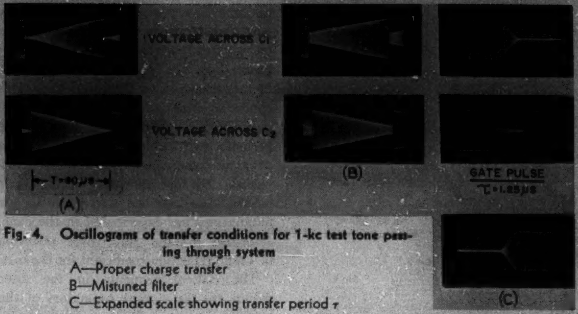


Fig. 4. Oscillograms of transfer conditions for 1-kc test tone passing through system
A—Proper charge transfer
B—Mistuned filter
C—Expanded scale showing transfer period τ

the modulation method to alleviate severe attenuation requirements on the low-pass filters.

A completely balanced transmission path was considered desirable to reduce extraneous noise and crosstalk coupling; however, the duplication of equipment and the increased complexity of the solid-state switching circuitry made this approach prohibitive. Single-ended multiplexing was selected and special care was given to its packaging: in particular to shielding, lead length, and the grounding system. Resonant transfer provides an additional asset in this respect since signal levels throughout the system are all at comparable levels. Signal level differences of 40 db or more are commonly encountered in straight sampling systems which further aggravate the noise and crosstalk problems.

Gate Design

The ultimate success of resonant-transfer multiplexing is largely dependent on the design of a gate circuit which has extremely low "on" resistance, a high-peak-signal current capacity, and a stable pulse width. In addition, the "off" impedance must be high to prevent highway loading and signal leakage into other time slots.

Transistor gates were investigated; however, they do not lend themselves readily to balanced modulator arrangements. A germanium diode quad was selected in the bridge configuration shown in Fig. 3(A) to produce the desired double polarity sampling. In the normal off condition the diode bridge is back-biased by the supply voltages of ±12 volts. The bridge is driven into conduction by a blocking oscillator which is triggered by the *MX* pulse assigned to any given line. The voltage conditions across the diodes are shown in Fig. 3(B) during the period of a single pulse.

During conduction the pulse current in the bridge must be several times larger than the peak signal current to be passed to prevent signal limiting. The peak signal current during the transfer time will be T/τ (or $80/1.25 = 64X$) larger than the steady-state signal current, i.e., all the charge stored in the frame period must be transferred in the short time of the sampling pulse to achieve zero-loss transmission. Obviously, the higher the impedance level of the signal appearing at the gate, the more relaxed will be the current requirement of the bridge. The maximum limit to which this can be carried is imposed by the back-bias of the bridge during the off period. Peak signals above this bias will tend to break through the gate, producing crosstalk in all channels. With a ± 12-volt battery supply, the impedance was set at 6 kilohms as a suitable compromise between minimum current and allowable peak voltage for a +3 dbm signal, the maximum signal level to be accommodated by the switchboard. Limiters are provided in the form of two Zener diodes back to back across the line-transformer secondary to insure that input signals do not exceed this peak voltage.

The peak transfer current for this signal level can be readily computed as approximately 50 ma (milliamperes), and the bridge pulse current must be in the order of 100 ma or more if limiting in the gate is to be avoided. At first glance, the instantaneous power-pulse requirements appear excessively large: peak current of 100 ma and an amplitude of approximately 26 volts (24 volts to overcome the bias plus the forward drop of two diodes) for a peak power of 2.6 watts. This condition is true if the diode bridge is to be driven from a straight pulse amplifier where the bias is supplied through a split secondary winding of an output transformer. However, in the blocking oscillator gate, the diode bridge replaces the clamping diode usually associated with this circuit to keep the transistor out of saturation. When the transistor enters the conduction state, both the collector and emitter tend toward ground thus removing the diode bridge bias during the on time. Consequently, the transistor supplies only the diode current and the magnetizing current to the transformer primary required to sustain the regenerative action. Short-circuiting the bias during sampling thus greatly reduces the pulse power required of the driving transistor stage.

Regeneration of the *MX* pulse by the

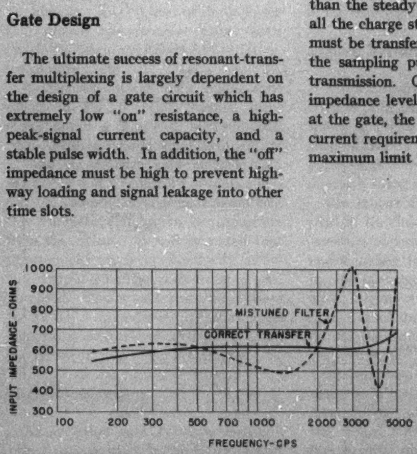


Fig. 5. Variations of line input impedance corresponding to transfer conditions of Fig. 4

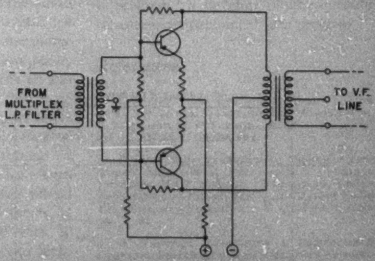


Fig. 6. Schematic diagram of balanced 10-db audio amplifier

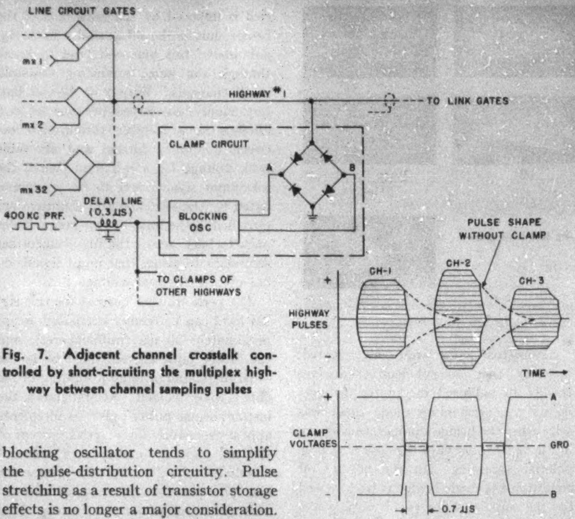

LINE CIRCUIT GATES
mx 1
mx 2
mx 32
HIGHWAY #1
TO LINK GATES
CLAMP CIRCUIT
A
B
DELAY LINE (0.3 µS)
400 KC PRF.
BLOCKING OSC
TO CLAMPS OF OTHER HIGHWAYS
PULSE SHAPE WITHOUT CLAMP
CH-1
CH-2
CH-3
+
HIGHWAY PULSES
-
TIME
+
A
CLAMP VOLTAGES
GRD
-
B
0.7 µS

Fig. 7. Adjacent channel crosstalk controlled by short-circuiting the multiplex highway between channel sampling pulses

blocking oscillator tends to simplify the pulse-distribution circuitry. Pulse stretching as a result of transistor storage effects is no longer a major consideration. Pulse width within the blocking oscillator is normally determined by the selection of the output transformer parameters. Variations in manufacture between units exceeded the desired specification of 1.25 ± 0.02 µsec; therefore the transformers were designed for a slightly wider pulse and the oscillators trimmed to the final value by padding the capacitor in the *R-C* network between the emitter and tertiary feedback winding.

Filter Design

The basis of the low-pass filter design for the multiplex path is in accordance with the insertion parameter theory. It is not intended to discuss here the intricacies of the specific design; however, several parameters should be mentioned which dictate the design approach.

The impedance level at the filters of 6 kilohms was previously determined by gate considerations. Filter synthesization on a steady-state matched-impedance basis does not provide entirely satisfactory transmission characteristics. This condition arises as a result of the time-varying load applied to the filter by the periodic switches. In addition to the departures encountered in expected transmission loss of the filters within their pass band as described in reference 3, wide fluctuations are experienced in the characteristic in-band impedance resulting in a high reflection coefficient for the line.

The critical elements in the filter design are the storage capacitors of the transfer circuit. These capacitors are required to perform two distinct functions: 1. capacitor C_1 must completely transfer its charge to C_2 during the sampling time, and 2. capacitor C_2 must completely distribute the received charge to the load during the remainder of the frame period $T-\tau$, so that its net charge is exactly zero at the start of the next transfer. A filter design which meets these conditions will insure satisfactory steady-state transmission and impedance characteristics of the multiplex path.

Fig. 4 consists of several oscillograms which illustrate these charge and discharge conditions. The oscillograph sweep is synchronized to the sampling frequency of 12.5 kc. Fig. 4(A) shows the voltage across capacitors C_1 and C_2 respectively for properly designed filters. Mistuning of the filters produces the results in Fig. 4(B). The amount of mistuning shown

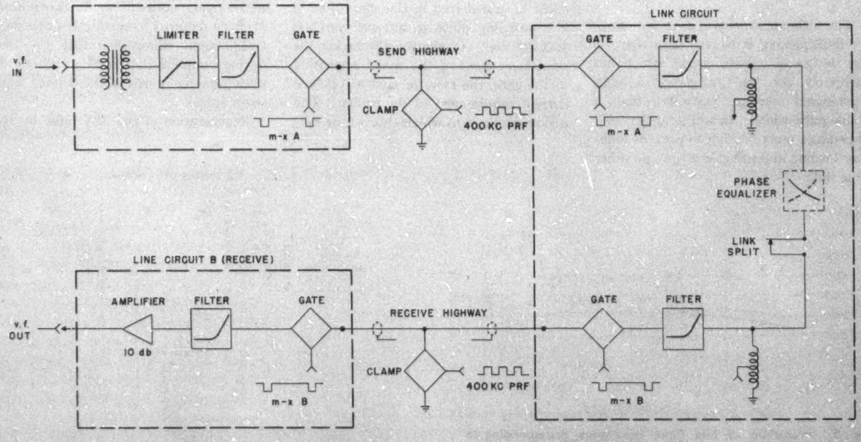

LINE CIRCUIT A (SEND)
v.f. IN
LIMITER
FILTER
GATE
m-x A
SEND HIGHWAY
CLAMP
400 KC PRF
LINK CIRCUIT
GATE
FILTER
m-x A
PHASE EQUALIZER
LINK SPLIT
LINE CIRCUIT B (RECEIVE)
v.f. OUT
AMPLIFIER
10 db
FILTER
GATE
m-x B
RECEIVE HIGHWAY
CLAMP
400 KC PRF
GATE
FILTER
m-x B

Fig. 8. Block diagram of complete transmission path through electronic switchboard

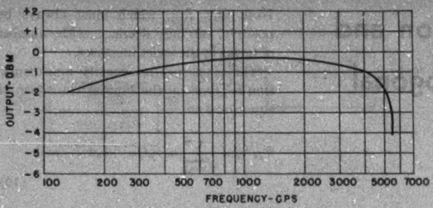


Fig. 9. Typical frequency response characteristic between voice-frequency drops of the switchboard

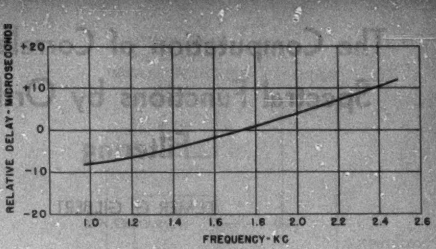


Fig. 10. Delay distortion over single multiplex connection

has very little effect on the frequency-response characteristic of the multiplex path; however, the input impedance variations of the line are quite large, as indicated in Fig. 5. Expanded scale photographs of the charge and discharge of the capacitors during the sampling time are shown in Fig. 4(C) with the diode bridge pulse as a reference.

Amplifier Design

To establish a call between line circuits within the switchboard, two steps of multiplexing are required: one between the calling line and a link, and the other between the link and the called line. Although the multiplexing losses are low in comparison to more conventional time-division systems, some amplification is required to achieve zero loss between voice drops. It was desirable to make up this loss at a single point to prevent duplication of equipment in both link and line circuits. An amplifier of 10-db power gain proved adequate to restore both multiplex losses and those due to filters, transformers, etc.

A schematic of the audio amplifier is shown in Fig. 6. Conventional techniques are used to provide bias stabilization. Both series and shunt feedback are utilized to stabilize gain against Beta variations and to control impedance levels; the balanced arrangement permitting the complete elimination of capacitors from the circuit. It will be noted that this amplifier can be operated with single-ended input since the bases are at ground potential. The input transformer is used to provide impedance transformation from the 6-kilohm level of the filters and a d-c return for the diode bridge gate.

Highway Clamping

As in other TDM systems, crosstalk varies with position in the frame period; the maximum crosstalk coupling occurring between adjacent channels in time as a result of pulse deformation by the transmission medium. In the electronic switchboard, distributed capacity of the highway is the major contributor to this form of crosstalk. To ease packaging requirements in the multiplex portion of the system, and to insure satisfactory crosstalk levels under varying highway conditions, a clamping circuit is used to discharge the highway in the guard time between sampling pulses. This circuit is another blocking oscillator gate with the diode bridge connected in shunt between the highway and ground as shown in Fig. 7. The blocking oscillator is triggered at a 400-kc pulse-repetition frequency from the master clock of the system and provides a pulse width of approximately 0.7 μsec. The pulse train is delayed by 0.3 μsec to prevent possible overlapping with the sampling pulses of the multiplex channels.

The short-circuiting action of the bridge between three active channels is indicated in the waveform sketches of Fig. 7.

System Performance

A block diagram of the complete transmission path through the switchboard is shown in Fig. 8. For simplicity, only a single direction of transmission is shown between two line circuits, the opposite direction being identical in circuitry.

The tapped chokes bridging the audio path in the link circuit are provided for insertion and monitoring of supervisory and signaling tones and to allow access for the operator's position. The chokes also provide d-c return for the diode bridge gates through the low-pass filters during the time that the audio path of the link is split for these operations.

Transmission measurements have been made on a large number of calls placed through the engineering test model of this board which is equipped with 30 line circuits, divided between two highways, and four complete links. A fully equipped service test model is presently undergoing tests. The frequency response of a typical connection through the switchboard is shown in Fig. 9. The total harmonic distortion for a 0-dbm test tone is less than 2% for any connection.

Demodulation products in the output, primarily composed of the side bands around the sampling frequency of 12.5 kc, average 32 to 34 db below test-tone output. Idle channel noise measures less than 15 dba, *FIA* weighted, for all connections. Maximum crosstalk between adjacent channels measures 55 db below the referenced output of 0 dbm.

The extremely low phase-delay distortion has proved to be a somewhat troublesome specification. To a large extent, the difficulty is centered around measurement techniques of sufficient accuracy to determine the existing conditions. Repeatable measurements of phase-angle differences to an accuracy of 0.01 degree is required. With the present filter design, a typical delay-distortion curve is shown in Fig. 10. These data indicate a difference of approximately 20 μsec for each multiplex hop or a total of 40 μsec per switchboard, which is approaching the practical limit obtainable from the filter design. Correction of this distortion to within the specification limits is expected to be accomplished in each link circuit by a phase equalizer presently being designed. Concurrently, a program to develop rapid, accurate measurement equipment is in progress.

References

1. Considerations Pertaining to the Design of an Electronic Telephone Switching System, B. Brightman, M. P. Tubinis. *AIEE Transactions*, see pp. 777–86 of this issue.

2. Steady-State Transmission Through Networks Containing Periodically Operated Switches, W. R. Bennett. *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., vol. CT-2, no. 1, 1955, pp. 17–21.

3. A Network Containing a Periodically Operated Switch Solved by Successive Approximations, C. A. Desoer. *Bell System Technical Journal*, New York, N. Y., vol. 36, 1957, pp. 1403–28.

The Computation of Correlation and Spectral Functions by Orthogonal Filtering

ELMER G. GILBERT
NONMEMBER AIEE

THE INCREASED consideration of random time functions in engineering and science has led to the development of various digital and analog techniques for the computation of autocorrelation and cross-correlation functions or their frequency domain equivalents, spectral density and cross-spectral density. Analog techniques, which are the subject of this paper, have been limited primarily to either the time domain[1–3] (employing the operations of time delay, multiplication, and averaging) or the frequency domain[4,5] (employing the operations of narrow-band filtering, multiplication, and averaging). A more general but little-known procedure, which was originally introduced by Lampard,[6] is to represent the autocorrelation or cross-correlation function by a finite series of predetermined orthogonal approximating functions and to determine coefficients in the series by the operations of orthogonal filtering, multiplication, and averaging. Advantages of the series method include analytic representations for both the correlation and spectral functions and the possible elimination of delay lines or narrow-band filters.

The purpose of this paper is to develop more fully the series method for computation of correlation and spectral functions. After reviewing the series representation and its determination, the following topics are treated: the relationship of time-domain and frequency-domain techniques to the series method; orthonormal approximating functions; analog computer circuits for orthonormal filtering; approximation errors due to a finite number of terms in the approximating series, computer component errors, and finite averaging time.

Series Representation and Its Determination

The cross-correlation function of $f(t)$ and $g(t)$ is defined by

$$\psi_{fg}(\tau)=\overline{f(t)g(t+\tau)}=\overline{f(t-\tau)g(t)} \tag{1}$$

where the bars denote averages over all time. It is assumed that f and g have finite power and that at least one of the spectral densities $\Psi_{ff}(j\omega)$ or $\Psi_{gg}(j\omega)$ is bounded. (Upper-case-letter functions of $j\omega$ are the Fourier transforms of corresponding lower-case-letter functions.) As shown in Appendix I, these are sufficient conditions to assure that:

$$\int_{-\infty}^{\infty} \psi_{fg}^2(\tau)d\tau<\infty \tag{2}$$

a condition which is required if the series representation is to hold. Since one of the spectral densities must be bounded, all periodic and constant components of that spectral must be excluded.

Let the series representation of ψ_{fg} be given by

$$\psi_{fg}^*=\sum_{n=1}^{N} [a_n\phi_n(\tau)+b_n\phi_n(-\tau)] \tag{3}$$

where the ϕ_n are orthonormal, i.e., they satisfy the conditions

$$\int_{-\infty}^{\infty} \phi_n(\tau)\phi_m(\tau)d\tau=0,\ n\neq m \atop \qquad\qquad =1,\ n=m \tag{4}$$

Orthonormality, rather than orthogonality, is stipulated for convenience. The second group of terms $b_n\phi_n(-\tau)$ in equation 3 is required because it is assumed that:

$$\phi_n(\tau)=0,\ \tau\leq 0 \tag{5}$$

Since equation 2 is satisfied, the integral square error

$$I=\int_{-\infty}^{\infty} (\psi_{fg}-\psi_{fg}^*)^2d\tau \tag{6}$$

is finite for finite a_n and b_n. Minimization of I in respect to the a_n and b_n gives the following results:

$$I_{\min}=\int_{-\infty}^{\infty} \psi_{fg}^2d\tau-\sum_{n=1}^{N} (a_n^2+b_n^2) \tag{7}$$

$$a_n=\int_{-\infty}^{\infty} \psi_{fg}(\tau)\phi_n(\tau)d\tau \tag{8}$$

and

$$b_n=\int_{-\infty}^{\infty} \psi_{fg}(\tau)\phi_n(-\tau)d\tau \tag{9}$$

Note that $I_{\min}$, which must always be positive, can only decrease as the number of terms in the series increases.

The cross-spectral density approximation is given by the Fourier transform of equation 3 as

$$\Psi_{fg}^*(j\omega)=\sum_{n=1}^{N}[a_n\Phi_n(j\omega)+b_n\Phi_n(-j\omega)] \tag{10}$$

Application of Parseval's theorem to $I_{\min}$ gives:

$$\begin{aligned} I_{\min}&=\frac{1}{2\pi}\int_{-\infty}^{\infty} |\Psi_{fg}(j\omega)-\Psi_{fg}^*(j\omega)|^2d\omega \\ &=\frac{1}{2\pi}\int_{-\infty}^{\infty} |\Psi_{fg}(j\omega)|^2d\omega- \\ &\qquad\sum_{n=1}^{N}(a_n^2+b_n^2) \end{aligned} \tag{11}$$

To obtain the representation for the autocorrelation function and spectral density of $f(t)$ it is only necessary to let $g=f$. Since $\psi_{ff}(-\tau)=\psi_{ff}(\tau)$, $b_n=a_n$ and equations 3 and 10 become simply:

$$\begin{aligned} \psi_{ff}^*(\tau)&=\sum_{n=1}^{N} a_n[\phi_n(\tau)+\phi_n(-\tau)] \\ &=\sum_{n=1}^{N} a_n\phi_n(|\tau|) \end{aligned} \tag{12}$$

and

$$\begin{aligned} \Psi_{ff}^*(j\omega)&=\sum_{n=1}^{N} a_n[\Phi_n(j\omega)+\Phi_n(-j\omega)] \\ &=2\sum_{n=1}^{N} a_n Re\Phi_n(j\omega) \end{aligned} \tag{13}$$

The procedure for evaluating the a_n in terms of $f(t)$ and $g(t)$ rather than $\psi_{fg}(\tau)$ is found by substituting equation 1 into equation 8 to obtain:

$$a_n=\int_{-\infty}^{\infty} \overline{f(t-\tau)g(t)}\phi_n(\tau)d\tau$$

Interchanging the order of the operations averaging and integration over τ gives:

$$a_n=\overline{g(t)h_n(t)} \tag{14}$$

where

$$h_n(t)=\int_{-\infty}^{\infty} f(t-\tau)\phi_n(\tau)d\tau \tag{15}$$

the response of a linear filter with input

Paper 59-1128, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted March 16, 1959; made available for printing July 23, 1959.

Elmer G. Gilbert is with the University of Michigan, Ann Arbor, Mich.

The author is grateful to F. J. Beutler and Edward O. Gilbert for helpful discussions during the preparation of the paper.

Vol.
Pt. 1
195

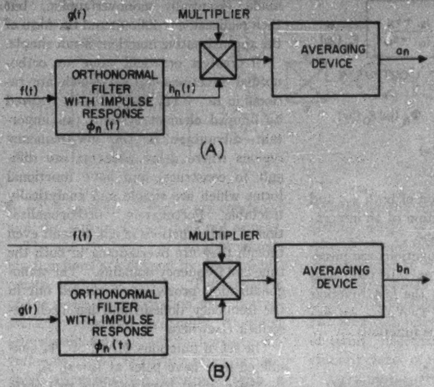

g(t)
MULTIPLIER
AVERAGING DEVICE
a_n
f(t)
ORTHONORMAL FILTER WITH IMPULSE RESPONSE $\phi_n(t)$
$h_n(t)$
(A)
f(t)
MULTIPLIER
AVERAGING DEVICE
b_n
g(t)
ORTHONORMAL FILTER WITH IMPULSE RESPONSE $\phi_n(t)$
(B)

Fig. 1. Physical systems for evaluation of a_n(A) and b_n(B)

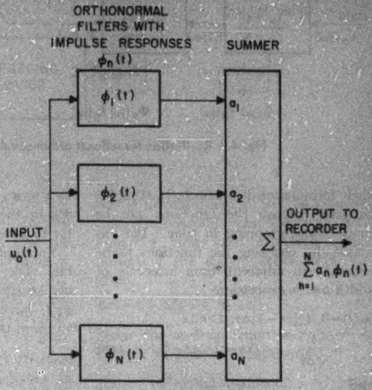

ORTHONORMAL FILTERS WITH IMPULSE RESPONSES $\phi_n(t)$
SUMMER
$\phi_1(t)$
$\phi_2(t)$
$\phi_N(t)$
a_1
a_2
a_N
INPUT $u_0(t)$
Σ
OUTPUT TO RECORDER
$\sum_{n=1}^{N} a_n\phi_n(t)$

Fig. 2. Method for recording positive τ portion of $\psi_{fg}{}^*$

$f(t)$ and impulse response $\phi_n(\tau)$. Since equation 5 holds, the impulse response $\phi_n(\tau)$ must be realizable. Fig. 1(A) shows the physical systems needed for evaluation 14. For computation of b_n a similar development gives the result in Fig. 1(B).

After the a_n and b_n have been computed the linear systems with impulse responses $\phi_n(\tau)$ can be employed as a function generator to machine plot the series representation $\psi_{fg}{}^*$. The scheme is shown in Fig. 2. The linear system synthesized has an impulse response $\sum_{n=1}^{N} a_n\ \phi_n(t)$. Thus when it is excited by an impulse $u_0(t)$, or a set of initial conditions producing the same effect as an impulse, an output equal to the positive τ portion of $\psi_{fg}{}^*$ is obtained for recording. The negative τ portion of $\psi^*{}_{fg}$ is recorded in exactly the same way by substituting the b_n for the a_n.

Generalizations of the foregoing procedures are possible. First of all, the approximating functions for negative τ need not be $\phi_1(-\tau), \ldots \phi_N(-\tau)$. Any number of other orthonormal functions, say $\bar{\phi}_1(-\tau), \ldots \bar{\phi}_M(-\tau)$, may be used by substituting $\bar{\phi}_n(-\tau)$ for $\phi(_n-\tau)$ in all the above results. Second, the approximating functions need not be orthogonal. When the ϕ_n are not orthogonal the computed results $\int_{-\infty}^{\infty} \psi_{fg}(\tau)\phi_n(\tau)d\tau$ and $\int_{-\infty}^{\infty} \psi_{fg}(\tau)\phi_n(-\tau)d\tau$ appear in $2N$ linear algebraic equations in $2N$ unknowns, a_n and b_n. While nonorthogonal functions are generally easier to realize as impulse responses, the solution of the linear equations for a_n and b_n is lengthy and tends to aggrevate the effect of errors in the computed $\int_{-\infty}^{\infty} \psi_{fg}(\tau)\phi_n(\tau)d\tau$ and $\int_{-\infty}^{\infty} \psi_{fg}(\tau)\phi_n(-\tau)d\tau$. Lastly, constrained or weighted integral-square-error approximations may be employed to diminish certain undesirable characteristics of the integral-square-error approximation.[7]

Relationship to Time-Domain and Frequency-Domain Techniques

In the usual time-domain computation of a correlation function the ϕ_n are made orthogonal (but not orthonormal) by causing the ϕ_n to be nonoverlapping functions of τ. Thus $\phi_n(t)=u_0(t-n\Delta\tau)$, the impulse response of a delay line with taps separated by $\Delta\tau$. By equations 8 and 9 it is seen that $a_n=\psi_{fg}(n\Delta\tau)$ and $b_n=\psi_{fg}(-n\Delta\tau)$. Actually, because of delay line errors in phase and gain, the $\phi_n(\tau)$ have finite width and height and only approximate $u_0(t-n\Delta\tau)$. Still, with a sufficiently good delay line the ϕ_n do not overlap appreciably and are essentially orthogonal.

A similar situation exists for the frequency-domain computation of spectral density. The functions ϕ_n are made orthogonal by causing their transforms $\Phi_n(j\omega)$ to be nonoverlapping in the frequency domain, i.e., by making the $\Phi_n(j\omega)$ be narrow band filters. (By Parseval's theorem:

$$\int_{-\infty}^{\infty}\phi_n(t)\phi_m(t)dt=\frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_n(j\omega)\times \Phi_m(-j\omega)d\omega$$

Hence functions orthogonal in the frequency domain are also orthogonal in the time domain.)

Orthonormal Approximating Functions

The primary factors governing the choice of orthonormal functions are functional form and physical realization. The functional form should yield series representations for ψ_{fg} and Ψ_{fg} which are accurate for reasonably small N and are easy to use; physical realization should be simple, accurate, and inexpensive.

When delay elements are allowed in the realization of the orthonormal functions

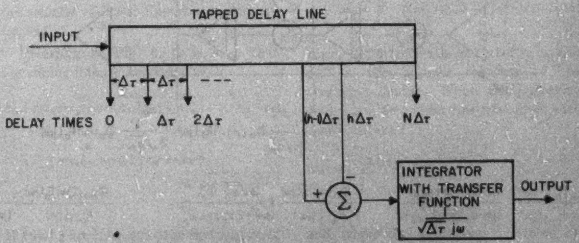

TAPPED DELAY LINE
INPUT
$\Delta\tau$
$\Delta\tau$
DELAY TIMES
0
$\Delta\tau$
$2\Delta\tau$
$(n-1)\Delta\tau$
$n\Delta\tau$
$N\Delta\tau$
+
−
Σ
INTEGRATOR WITH TRANSFER FUNCTION $\frac{1}{\sqrt{\Delta\tau}\ j\omega}$
OUTPUT

Fig. 3. Realization for orthonormal function $\phi_n(\tau)$

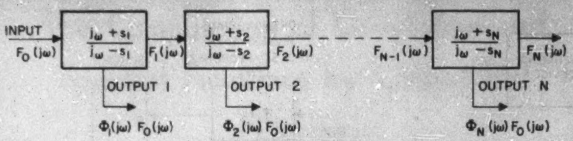

INPUT
$F_0(j\omega)$
$\frac{j\omega+s_1}{j\omega-s_1}$
$F_1(j\omega)$
$\frac{j\omega+s_2}{j\omega-s_2}$
$F_2(j\omega)$
$F_{N-1}(j\omega)$
$\frac{j\omega+s_N}{j\omega-s_N}$
$F_N(j\omega)$
OUTPUT 1
OUTPUT 2
OUTPUT N
$\Phi_1(j\omega)F_0(j\omega)$
$\Phi_2(j\omega)F_0(j\omega)$
$\Phi_N(j\omega)F_0(j\omega)$

Fig. 4. Realization for rational orthonormal function $\Phi_n(j\omega)$

their formulation is particularly straightforward because the functions can be made nonoverlapping in time. For example, by choosing the functions to be constant in adjacent time intervals of width $\Delta\tau$ the expressions

$$\begin{aligned}\phi_n(t)&=0,\ t\leq(n-1)\Delta\tau,\ t>n\Delta\tau\\&=\frac{1}{\sqrt{\Delta\tau}},\ (n-1)\Delta\tau<t\leq n\Delta\tau\end{aligned}\tag{16}$$

and

$$\Phi_n(j\omega)=\frac{\epsilon^{-j\omega(n-1)\Delta\tau}-\epsilon^{-j\omega n\Delta\tau}}{\sqrt{\Delta\tau}\,(j\omega)}\tag{17}$$

result. Here a stepwise representation of $\psi_{fg}(\tau)$ is achieved. The frequency-domain approximation is a harmonic series of sines and cosines (of fundamental period $\omega=2\pi/\Delta\tau$) weighted by $1/\omega$. A physical realization of the functions is shown in Fig. 3. In practice the integrator might be an R-C low-pass filter with time constant much greater than $\Delta\tau$. Thus it is possible to go from a point by point determination of $\psi_{fg}(at\ \tau=n\Delta\tau)$ to a series representation of both ψ_{fg} and Ψ_{fg} by the simple addition of an integrator to a tapped delay line.

Variations of these functions are possible. The adjacent time intervals need not be equal, nor does the last function $\phi_N(t)$ have to be zero $t>N\Delta\tau$. An example of the latter is the function:

$$\begin{aligned}\phi_N(t)&=0,\ t\leq(N-1)\Delta\tau\\&=\sqrt{2a}\,\epsilon^{a(n-1)\Delta\tau}\,\epsilon^{-at},\ t>(N-1)\Delta\tau\end{aligned}\tag{18}$$

Still more varied functions may be realized with delay elements if the nonoverlapping requirement is not enforced. But then, of course, orthogonalization is not so easily carried out.

Attempts to form orthogonal functions which are nonoverlapping in the frequency domain encounter theoretical limitations because of realizability considerations. (The Paley-Weiner theorem shows that $\Phi_n(j\omega)$ is not realizable if $\Phi_n(j\omega)=0$ in any finite frequency range.) Band-pass filters with extremely sharp cutoff can be made essentially nonoverlapping, but their realization is difficult and the form of the approximating functions is not simple.

Of greatest practical value are orthonormal functions $\Phi_n(j\omega)$ which are rational in $j\omega$. They may be easily realized by lumped element systems (an important advantage in the low-frequency regions where delay elements are difficult to construct) and have functional forms which are simple and analytically tractable. Fortunately, orthogonalization of such functions is not difficult even though they are overlapping in both the time and frequency domains. The orthogonalization procedure is carried out in the frequency domain and has been described elsewhere.[7]

The set of functions $\Phi_1, \Phi_2, \ldots \Phi_n$ considered here have poles at $j\omega=s_1, s_2, \ldots s_n$ (the s_n must have negative real parts since $\int_{-\infty}^{\infty}\phi_n^2dt=1<\infty$). When s_n is real with $s_n=-\alpha_n$:

$$\Phi_n=\frac{(j\omega+s_1)\ldots(j\omega+s_{n-1})}{(j\omega-s_1)\ldots(j\omega-s_{n-1})}\,\frac{\sqrt{2\alpha_n}}{j\omega+\alpha_n}\tag{19}$$

When s_n and s_{n+1} are complex conjugates with real parts $-\alpha_n$ and magnitudes β_n,

$$\Phi_n=\frac{(j\omega+s_1)\ldots(j\omega+s_{n-1})}{(j\omega-s_1)\ldots(j\omega-s_{n-1})}\times\frac{2\beta_n\sqrt{\alpha_n}}{(j\omega)^2+2\alpha_nj\omega+\beta_n^2}\tag{20}$$

and

$$\Phi_{n+1}=\frac{(j\omega+s)\ldots(j\omega+s_{1n-1})}{(j\omega-s_1)\ldots(j\omega-s_{n-1})}\times\frac{2\sqrt{\alpha_n}j\omega}{(j\omega)^2+2\alpha_nj\omega+\beta_n^2}\tag{21}$$

The corresponding time functions $\phi_n(t)$ are not written so easily and therefore are not given here. When all the s_n are real and equal, the functions ϕ_n are proportional to the Laguerre functions utilized by Lampard.

A realization of the foregoing functions is shown in Fig. 4. Since the higher number functions include the dynamic characteristics of the lower number functions, the cascade arrangement provides minimum complexity. Operational amplifier circuits for the sections shown in Fig. 4 are given in Fig. 5. Except for the normalization factors $+1/2\sqrt{\alpha_n}$ and $-1/\sqrt{2\alpha_n}$ these sections may be fitted directly into Fig. 4. For the functions defined by equations 20 and 21 the n^{th} and

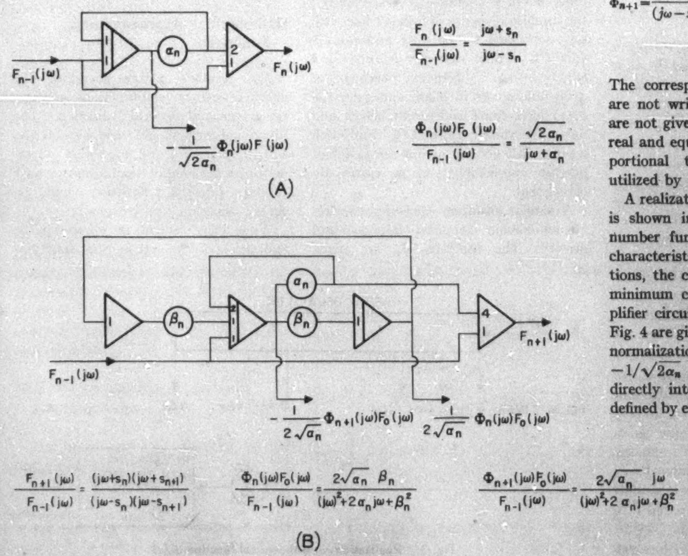

$F_{n-1}(j\omega)$
1
1
α_n
2
1
$F_n(j\omega)$
$-\frac{1}{\sqrt{2\alpha_n}}\Phi_n(j\omega)F(j\omega)$
$\frac{F_n(j\omega)}{F_{n-1}(j\omega)}=\frac{j\omega+s_n}{j\omega-s_n}$
$\frac{\Phi_n(j\omega)F_0(j\omega)}{F_{n-1}(j\omega)}=\frac{\sqrt{2\alpha_n}}{j\omega+\alpha_n}$
(A)
1
β_n
2
1
1
α_n
β_n
1
4
1
$F_{n+1}(j\omega)$
$F_{n-1}(j\omega)$
$-\frac{1}{2\sqrt{\alpha_n}}\Phi_{n+1}(j\omega)F_0(j\omega)$
$\frac{1}{2\sqrt{\alpha_n}}\Phi_n(j\omega)F_0(j\omega)$
$\frac{F_{n+1}(j\omega)}{F_{n-1}(j\omega)}=\frac{(j\omega+s_n)(j\omega+s_{n+1})}{(j\omega-s_n)(j\omega-s_{n+1})}$
$\frac{\Phi_n(j\omega)F_0(j\omega)}{F_{n-1}(j\omega)}=\frac{2\sqrt{\alpha_n}\ \beta_n}{(j\omega)^2+2\alpha_nj\omega+\beta_n^2}$
$\frac{\Phi_{n+1}(j\omega)F_0(j\omega)}{F_{n-1}(j\omega)}=\frac{2\sqrt{\alpha_n}\ j\omega}{(j\omega)^2+2\alpha_nj\omega+\beta_n^2}$
(B)

Fig. 5. Operational amplifier circuits for sections shown in Fig. 4

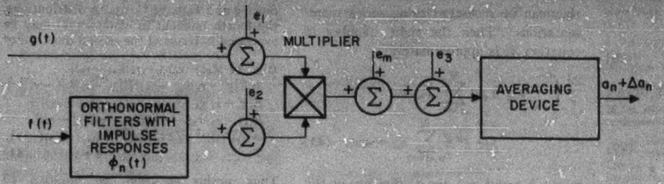


Fig. 6. Diagram showing component errors

$(n+1)$th sections in Fig. 4 are combined and the output $F_n(j\omega)$ does not appear.

The singularities s_n of the approximating functions are defined by the potentiometer settings α_n and β_n. Thus the form of the series approximation may be conveniently changed by simple computer adjustment without any preliminary calculations. This flexibility often permits good approximation accuracy with a rather small N, an advantage which the Laguerre functions do not have.

Approximation Errors

Three distinct sources of approximation error exist: 1. the finite number of terms in the approximating series, 2. deviations in the a_n and b_n due to computer component errors, and 3. statistical deviations in the a_n and b_n due to finite averaging time. The first error source yields an error magnitude measured by $I_{\min}$ in equation 7. If the deviations in a_n and b_n caused by the last two error sources are denoted by Δa_n and Δb_n, then it can be shown that the total integral square approximation error is:[7]

$$I=I_{\min}+\Delta I=I_{\min}+\sum_{n=1}^{N}(\Delta a_n^2+\Delta b_n^2) \tag{22}$$

Thus ΔI represents the integral square error due to sources *2* and *3*. Practically there is no point in striving to make either $I_{\min}<<\Delta I$ or $\Delta I<<I_{\min}$. The three error sources will now be examined in detail.

Errors Due to Finite N

The N required for a given approximation accuracy varies widely with the choice of the approximating functions ϕ_n. Often, several of the rational functions given in equations 19, 20, and 21 will give a very satisfactory representation if they are chosen with care. However, since ψ_{fg} is initially unknown, careful choice of the functions is usually not feasible and large number of ϕ_n with varied characteristics are needed. The necessary variation of characteristics can be achieved with rational functions by distributing the points of singularity s_n throughout the complex plane in the frequency range of importance. In any case, N should not differ greatly from the N demanded in the more usual time or frequency-domain computations of ψ_{fg}. Once a reasonable representation of ψ_{fg} has been computed using a large number of terms, inspection may lead to a more efficient choice of the ϕ_n.

An additional complication arises in evaluating the goodness of the approximation since $\int_{-\infty}^{\infty}\psi_{fg}^2 d\tau$ is not known and $I_{\min}$ cannot be computed. However, comparative calculations are possible, since $\sum_{n=1}^{N}(a_n^2+b_n^2)$ can be computed and becomes larger as approximation accuracy improves. Furthermore, a check on the approximation error at $\tau=0$ is readily made since $\psi_{fg}(0)=\overline{f(t)g(t)}$ is easily computed. This check is valuable when the rational approximating functions are utilized because the discontinuities in many of the $\phi_n(\tau)$ at $\tau=0$ often cause the approximation error to be greatest at $\tau=0$. Of course, the accuracy of these checks is limited by computer component errors and finite averaging time.

Computer Component Errors

The component errors considered in this section are indicated in Fig. 6. The quantities e_1, e_2, and e_3 are constants representing the effects of offsets in the computer circuitry. The multiplier static error is e_m and is a function of the two inputs to the multiplier. Dynamic errors in the multiplier and realization of the ϕ_n are usually of lesser importance and are neglected.

Referring to Fig. 6, it is seen that (assuming infinite averaging time):

$$\begin{aligned} a_n+\Delta a_n &= \overline{(g+e_1)(h_n+e_2)}+\overline{e_m(g+e_1,\ h_n+e_2)} \\ &\quad +\overline{e_3} \\ &= e_1\overline{h_n}+e_2\overline{g}+\overline{gh_n}+e_1e_2+ \\ &\quad \overline{e_m(g+e_1,\ h_n+e_2)} \\ &\quad +e_3 \end{aligned} \tag{23}$$

If $\bar{f}$ and $\bar{g}$ are both zero, the first two error terms in equation 23 vanish. Because this is advantageous it will be assumed that the constant components of both f and g have been extracted. Then neglecting the second order term, e_1e_2 gives:

$$\Delta a_n=e_3+\overline{e_m(g+e_1,\ h_n+e_2)} \tag{24}$$

Since the average multiplier error is difficult to compute, the bound

$$|\Delta a_n|\le|e_3|+|e_m|\ \text{maximum} \tag{25}$$

is more useful. The same result applies to Δb_n. From the above it can be seen that multiplier error and integrator drift are the important component errors.

To keep the effect of these errors as small as possible, the levels of f, g, and h_n should be scaled as large as possible, observing saturation limitations. In this regard servo multipliers are severely limited, even when their linear-system bandwidth is sufficiently large. Velocity and acceleration limits can be encountered long before full-scale amplitudes are reached.

Statistical Errors Due to Finite Averaging Time

The time averaging of $g(t)h_n(t)$ which yields a_n must necessarily be carried out by a low-pass filter or integrator. In either situation a statistical error results because the average is not carried out over all time.[5,9] To derive the results that follow it is assumed that f and g belong to stationary and ergodic random processes.

If the average of $g(t)h_n(t)$ is measured at time t the result can be expressed by:

$$a_n+\Delta a_n=\int_0^{\infty} g(t-\tau)h_n(t-\tau)w(\tau)d\tau \tag{26}$$

where Δa_n is the statistical error in a_n, and $w(t)$ is the impulse response of the averaging device. When the averaging device is a first-order low-pass filter with time constant T,

$$\begin{aligned} w(t) &= \frac{1}{T}e^{-t/T}, & t>0 \\ &= 0, & t\le 0 \end{aligned} \tag{27}$$

and when the averaging device is an integrator with averaging time T,

$$w(t)=\frac{1}{T}, \quad 0<t\leq T$$
$$=0, \quad t\leq 0, \quad t>T \tag{28}$$

The proper gain of the averaging device is achieved by requiring

$$\int_0^\infty w(t)dt=1 \tag{29}$$

Because $g(t)$ and $f(t)$ are stationary and ergodic, the expected value of $a_n+\Delta a_n$ is

$$\begin{aligned} E[a_n+\Delta a_n]&=a_n+E[\Delta a_n]\\ &=\int_0^\infty E[g(t-\tau)h_n(t-\tau)]w(\tau)d\tau\\ &=E[gh_n]\int_0^\infty w(\tau)d\tau\\ &=E[gh_n]=\overline{g(t)h_n(t)}=a_n \end{aligned} \tag{30}$$

Thus $E\ [\Delta a_n]=0$ and the estimate of a_n is unbiased. In an analogous fashion the same result can be developed for Δb_n.

In order to derive reasonably simple expressions for the variance of Δa_n it is necessary to further assume that g and f are Gaussian. For processes which are not Gaussian the results presented here will at least serve as an approximate guide. In Appendix II it is shown that:

$$\begin{aligned} E\left[\sum_{n=1}^N \Delta a_n^2+\Delta b_n^2\right]&=E[\Delta I]\\ &=\frac{k}{\pi}\int_{-\infty}^\infty\left[\Psi_{gg}\Psi_{ff}\sum_{n=1}^N|\Phi_n|^2+\Psi_{fg}^2\times\right.\\ &\qquad\left.\sum_{n=1}^N Re\Phi_n^2\right]d\omega \end{aligned} \tag{31}$$

where

$$\begin{aligned} k&=\int_0^\infty w(t)^2dt\\ &=\frac{1}{T}\text{ for equation 27}\\ &=\frac{1}{2T}\text{ for equation 28} \end{aligned} \tag{32}$$

Thus the average increment of integral-square-approximation error is inversely proportional to the averaging time.

Since equation 31 is somewhat too complicated to use for simple estimates of ΔI, the following upper bound on $E[\Delta I]$ is derived in Appendix III:

$$E[\Delta I]\leq\frac{2k}{\pi}\int_{-\infty}^\infty \Psi_{gg}\Psi_{ff}\sum_{n=1}^N|\phi_n|^2d\omega \tag{33}$$

Once $\sum_{n=1}^N |\Phi_n|^2$ has been evaluated for a given set of approximating functions, the above integral can be quickly estimated to determine a suitable averaging time. If Ψ_{gg} and Ψ_{ff} are not known initially, they can be measured using series representations. Then the right member of equation 33 is approximately:

$$\begin{aligned} \frac{2k}{\pi}\int_{-\infty}^\infty \Psi_{gg}{}^*\Psi_{ff}{}^*\sum_{n=1}^N|\Phi_n|^2d\omega\\ =k\sum_{n,m=1}^N a_{gn}a_{fm}c_{nm} \end{aligned} \tag{34}$$

where a_{gn} and a_{fm} are coefficients in the series $\Psi_{gg}{}^*$ and $\Psi_{ff}{}^*$ and

$$c_{nm}=\frac{8}{\pi}\int_{-\infty}^\infty Re\Phi_n Re\Phi_m\sum_{j=1}^N|\Phi_j|^2d\omega \tag{35}$$

All of these results can be applied to autocorrelation function measurements by replacing the g subscripts by f subscripts.

Appendix I

Let λ be an arbitrary real parameter. Then the spectral density of the time function $\lambda f+g$ must be nonnegative. Thus:

$$0\leq\Psi_{(\lambda f+g)(\lambda f+g)}=\lambda^2\Psi_{ff}+\lambda\Psi_{fg}+\lambda\Psi_{gf}+\Psi_{gg}$$

But $\Psi_{gf}=\text{conj }\Psi_{fg}$, so:

$$\lambda^2\Psi_{ff}+2\lambda Re\Psi_{fg}+\Psi_{gg}\geq 0$$

Since the result must hold for all values of λ the discriminant must be negative or zero and

$$4(Re\Psi_{fg})^2-4\Psi_{ff}\Psi_{gg}\leq 0$$

or

$$(Re\Psi_{fg})^2\leq\Psi_{ff}\Psi_{gg} \tag{36}$$

Similar consideration of the spectra density of $\lambda f+jg$ yields the result:

$$(Im\Psi_{fg})^2\leq\Psi_{ff}\Psi_{gg} \tag{37}$$

Addition of equations 36 and 37 gives:

$$|\Psi_{fg}|^2\leq 2\Psi_{ff}\Psi_{gg} \tag{38}$$

Now if one of the spectral densities is bounded (say $\Psi_{ff}\leq M$), then:

$$\begin{aligned} \int_{-\infty}^\infty \psi_{fg}^2(\tau)d\tau&=\frac{1}{2\pi}\int_{-\infty}^\infty|\Psi_{fg}|^2d\omega\leq\\ &\frac{1}{\pi}\int_{-\infty}^\infty\Psi_{gg}\Psi_{ff}d\omega\leq\frac{M}{\pi}\int_{-\infty}^\infty\Psi_{gg}d\omega \end{aligned}$$

and

$$\int_{-\infty}^\infty \psi_{fg}^2(\tau)d\tau\leq M2\phi_{gg}(0)<\infty \tag{39}$$

This is the desired result.

Appendix II

The variance of Δa_n can be written as

$$E[\Delta a_n^2]=E[(a_n+\Delta a_n)^2]-a_n^2$$

Substituting $a_n+\Delta a_n$ from equation 26 and for convenience letting $t=0$ gives:

$$\begin{aligned} E[\Delta a_n^2]=\int_0^\infty\int_0^\infty E[g(-\tau_1)h_n(-\tau_1)\times\\ g(-\tau_2)h_n(-\tau_2)]w(\tau_1)w(\tau_2)d\tau_1d\tau_2-a_n^2 \end{aligned} \tag{40}$$

Since g and h_n have Gaussian distributions, the fourth moment in equation 40 may be expressed in terms of the second moments:[10]

$$\begin{aligned} &E[g(-\tau_1)h_n(-\tau_1)g(-\tau_2)h_n(-\tau_2)]\\ &\quad=E[g(-\tau_1)h_n(-\tau_1)]E[g(-\tau_2)h_n(-\tau_2)]+\\ &\quad E[g(-\tau_1)g(-\tau_2)E[h_n(-\tau_1)h_n(-\tau_2)]+\\ &\quad E[g(-\tau_1)h_n(-\tau_2)]E[g(-\tau_2)h_n(-\tau_1)]\\ &\quad=\psi_{gh}^2(0)+\psi_{gg}(\tau)\psi_{hh}(\tau)+\psi_{hg}(\tau)\psi_{hg}(-\tau),\\ &\qquad\tau=\tau_1-\tau_2 \end{aligned} \tag{41}$$

Thus, noting equation 29, equation 40 becomes:

$$E(\Delta a_n^2)=\int_0^\infty\int_0^\infty\theta(\tau_1-\tau_2)w(\tau_1)\times w(\tau_2)d\tau_1d\tau_2 \tag{42}$$

where

$$\theta(\tau)=\theta(-\tau)=\psi_{gg}(\tau)\psi_{hh}(\tau)+\psi_{hg}(\tau)\psi_{hg}(-\tau) \tag{43}$$

By defining

$$\phi_{ww}(\tau)=\int_0^\infty w(\tau_1)w(\tau+\tau_1)d\tau_1 \tag{44}$$

equation 42 can be written as:

$$E[\Delta a_n^2]=\int_{-\infty}^\infty\theta(\tau)\theta_{ww}(\tau)d\tau\leq k\int_{-\infty}^\infty\theta(\tau)d\tau \tag{45}$$

where $k=\phi_{ww}(0)$ is the upper bound of $\phi_{ww}(\tau)$. Since the averaging times are usually very long compared to values of $|\tau|$ where $\psi_{ff}(\tau)$, $\psi_{gg}(\tau)$, $\psi_{fg}(\tau)$, and consequently $\theta(\tau)$ have appreciable magnitude, $E[\Delta a_n^2]$ approaches closely the upper bound given in equation 45. Thus:

$$E[\Delta a_n^2]\cong k\int_{-\infty}^\infty\theta(\tau)d\tau \tag{46}$$

Application of Parseval's theorem to equation 46 gives:

$$E[\Delta a_n^2]\cong\frac{k}{2\pi}\int_{-\infty}^\infty[\Psi_{gg}(j\omega)\Psi_{hh}(-j\omega)+\Psi_{hg}^2(j\omega)]d\omega \tag{47}$$

But from equation 15

$$\Psi_{hh}(-j\omega)=\Psi_{hh}(j\omega)=\Psi_{ff}(j\omega)|\Phi_n(j\omega)|^2$$

and

$$\Psi_{hg}(j\omega)=\Psi_{fg}(j\omega)\Phi_n(-j\omega)$$

Thus

$$E[\Delta a_n^2]\cong\frac{k}{2\pi}\int_{-\infty}^\infty[\Psi_{gg}\Psi_{ff}|\Phi_n|^2+(\Psi_{fg}(j\omega)\Phi_n(-j\omega))^2]d\omega \tag{48}$$

Similarly,

$$\begin{aligned} E[\Delta b_n^2]&\cong\frac{k}{2\pi}\int_{-\infty}^\infty[\Psi_{gg}\Psi_{ff}|\Phi_n|^2+(\Psi_{fg}(-j\omega)\Phi_n(-j\omega))^2]d\omega\\ &\cong\frac{k}{2\pi}\int_{-\infty}^\infty[\Psi_{gg}\Psi_{ff}|\Phi_n|^2+(\Psi_{fg}(j\omega)\Phi_n(j\omega))^2]d\omega \end{aligned} \tag{49}$$

Summation of the terms in equations 48 and 49 gives equation 31.

Appendix III

Equation 47 can be written as:

$$E[\Delta a_n^2]=\frac{k}{2\pi}\int_{-\infty}^{\infty}[\Psi_{gg}(j\omega)\Psi_{hh}(j\omega)+$$

$$Re\Psi_{hg}^2(j\omega)+jIm\Psi_{hg}^2(j\omega)]d\omega$$

Since $E[\Delta a_n^2]$ is real the integral of $Im\Psi_{hg}^2$ must vanish. Hence:

$$E[\Delta a_n^2]=\frac{k}{2\pi}\int_{-\infty}^{\infty}[\Psi_{gg}\Psi_{hh}+Re\Psi_{hg}^2]d\omega$$

But $Re\Psi_{hg}^2(j\omega)\leq(Re\Psi_{hg}(j\omega))^2$. Hence:

$$E[\Delta a_n^2]\leq\frac{k}{2\pi}\int_{-\infty}^{\infty}[\Psi_{gg}\Psi_{hh}+(Re\Psi_{hg})^2]d\omega$$

Application of inequality 36 with f replaced by h gives:

$$E[\Delta a_n^2]\leq\frac{2k}{2\pi}\int_{-\infty}^{\infty}\Psi_{gg}\Psi_{hh}d\omega$$

$$=\frac{k}{\pi}\int_{-\infty}^{\infty}\Psi_{gg}\Psi_{ff}\Phi_n^2d\omega \quad (50)$$

Similarly, the same upper bound applies to $E[\Delta b_n^2]$.

The summation

$$\sum_{n=1}^{N}(E[\Delta a_n^2]+E[\Delta b_n^2])$$

$$=E\left[\sum_{n=1}^{N}\Delta a_n^2+\Delta b_n^2\right]$$

therefore gives equation 33.

All of the foregoing results can be applied to autocorrelation function measurements by replacing the g subscripts by f subscripts.

References

1. A DEVICE FOR COMPUTING CORRELATION FUNCTIONS, A. E. Hastings, J. E. Meade. *Review of Scientific Instruments*, New York, N. Y., vol. 23, 1952, p. 347.

2. A FIVE CHANNEL ELECTRONIC ANALOG CORRELATOR, M. J. Levin, J. F. Reintjes. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 8, 1952, p. 647.

3. A HIGH-SPEED CORRELATOR, H. Bell, V. C. Rideout. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., vol. PGEC-3, no. 2, June 1954, p. 70.

4. *AFOSR TN 58-338*, A. Huberstrich, F. R. Hama. USAF Office of Scientific Research, Washington, D. C., 1958.

5. TECHNIQUE FOR MEASUREMENT OF CROSS-SPECTRAL DENSITY OF TWO RANDOM FUNCTIONS, M. S. Uberoi, E. G. Gilbert. *Review of Scientific Instruments*, vol. 30, Mar. 1959, p. 176.

6. A METHOD OF DETERMINING CORRELATION FUNCTIONS OF STATIONARY TIME SERIES, D. G. Lampard. *Proceedings*, Institution of Electrical Engineers, London, England, pt. C, no. 1, Mar. 1955, p. 35.

7. LINEAR SYSTEM APPROXIMATION BY DIFFERENTIAL ANALYZER SIMULATION OF ORTHONORMAL APPROXIMATING FUNCTIONS, E. G. Gilbert. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, vol. PGEC-8, no. 2, June 1959.

8. STATISTICAL ERRORS IN MEASUREMENTS ON RANDOM TIME FUNCTIONS, W. B. Davenport, Jr., R. A. Johnson, D. Middleton. *Journal of Applied Physics*, New York, N. Y., vol. 23, Apr. 1952, p. 377.

9. THE MEASUREMENT OF POWER SPECTRA FROM THE POINT OF VIEW OF COMMUNICATIONS ENGINEERING, R. B. Blackman, J. W. Tukey. *Bell System Technical Journal*, New York, N. Y., vol. 37, 1958, PART I, p. 185; PART II, p. 485.

10. RANDOM PROCESSES IN AUTOMATIC CONTROL (book), J. H. Laning, Jr., R. H. Battin. McGraw-Hill Book Company, Inc., New York, N. Y., 1956.

A Small Automatic Teletypewriter Switching System

E. R. SHIMMIN
NONMEMBER AIEE

R. A. VANDERLIPPE
NONMEMBER AIEE

A. L. WHITMAN
MEMBER AIEE

TWO teletypewriter selective calling line systems, the TADS (Teletypewriter Automatic Dispatch System) system[1] and the *83B1* system,[2] have been described in recent papers. These systems are for single multistation lines; those in service at the present time have an average of about ten stations on a line although, theoretically, if the traffic offered per station is low, there may be as many as 40 stations on a line. As soon as the required number of stations exceeds that which can be efficiently handled on one line, other lines are required, with the need for transferring messages from one line to another.

For simple networks of not more than three lines, the transferring can be done efficiently and economically by an existing automatic relay arrangement using reperforator-transmitters with intermediate punched tape as the switching device. For more than three lines such a switching device becomes increasingly more complicated and prohibitively expensive.

To fill the need for interconnecting from four to nine selective calling lines a new automatic switching unit, known as the *84A1* teletypewriter switching unit, was developed. In the interests of simplicity and economy, the requirements for the system were set with certain recognized limitations. As a result, the capacity of the switching unit to handle crossoffice traffic efficiently is limited. There is no fixed upper limit of capacity; what happens if the system becomes overloaded is that the messages are subject to longer delays in delivery. The plan for the *84A1* system was laid out on the basic assumption that it would be used in situations where the principal communities of interest would be along the individual selective calling lines, and that only a relatively small amount of traffic would be exchanged between lines. Quantitatively, about 60% of the total traffic on a particular line should be intraline, with only 40% interline.

This will indicate why the system has been called a "small" system. It is in no sense competitive, so far as traffic-handling capacity is concerned, with the larger and more elaborate automatic switching systems of the *81* type.[3] However, it is believed that it will fill a significant need for businesses which do not need or cannot afford the larger systems. In addition, where the customer has a need to terminate some half-duplex multistation circuits, these circuits could terminate on an *84A1* switching unit. The transfer of traffic between the two systems could be effected manually by the torn tape method or by an automatic interconnection now under consideration.

A significant difference between the *84* and *81* systems, is that the lines of the *84* system operate half duplex (HDX), while the lines of the *81* system are full duplex (FDX). Half-duplex operation permits free interchange of traffic between stations on one line without the intervention of the switching center. If a substantial proportion of the traffic is of this kind, HDX operation relieves the switching center of a large part of its load.

The multiple-address capacity of the switching unit is limited only by the total number of stations on the lines switched by the unit. Since the director processes each address separately, it will continue to set up crossoffice paths

---

Paper 59-1105, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11-16, 1959. Manuscript submitted April 22, 1959; made available for printing July 15, 1959.

E. R. SHIMMIN is with the Pacific Telephone and Telegraph Company, San Francisco, Calif.; R. A. VANDERLIPPE and A. L. WHITMAN are with Bell Telephone Laboratories, Inc., New York, N. Y.

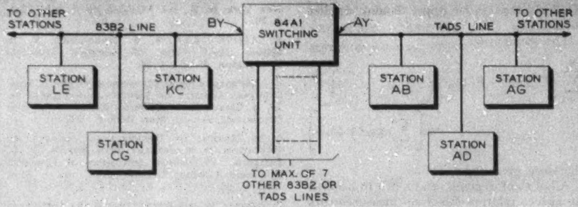


Fig. 1. Typical circuit layout of 84A1 switching system

for as many addresses as the originating operator has punched in the tape.

Group codes may be used in a limited way. A single group code will call in stations on only one line. The switching unit will not set up connections to two or more lines on one code. A group code, reaching its destination line, will connect, on that line, all of the stations to which the code applies. As many group codes can be used as desired, provided that each of them calls stations on only one line.

The unit will not handle a single broadcast code for all lines or all stations of a system.

Typical Circuit Layout

Fig. 1 shows a typical circuit layout of an *84A1* system, with the connecting selective calling lines, which may be of either the TADS[1] or the *83B2*[2] variety. Parenthetically, the *83B2* system is a slightly modified version of the *83B1* system described in reference 2, the changes being principally to increase flexibility in the assignment of call directing codes (CDCS) and to simplify and improve interline switching arrangements.

It is recalled that the basic features of the selective calling lines are HDX operation, at 60, 75, or 100 words per minute. The selective calling features are principally the automatic starting of the station tape transmitters, and automatic selection of the desired receiving stations, so that the message is received at only the stations to which it is addressed.

The switching center, on the line side, operates just like any other station on the line except that for convenience of circuit design the selective calling control station, with the transmitter start circuit, is always located at the switching center. The connection between the lines and the switching unit is made, as Fig. 2 shows, with two reperforator-transmitters for each line, one for incoming traffic (ICL RT) and the other for outgoing traffic (OGL RT). The reperforator of the ICL RT is called in by its CDC just like any other receiver on the line. The tape transmitter of the OGL RT is started in its turn (when it is waiting with a message which has been switched across the office for the line) by a transmitter start code just like any other transmitter on the line.

One difference is that the reperforator-transmitters handle only traffic to and from the switching unit. If the location of the switching center is also a business office, it will be necessary to provide one or more extra line stations there. To relieve the crossoffice load as far as possible, these local stations should be on the lines which have the most traffic for the switching center location. The "local outlet" of the switching center is provided by one or more line stations.

The *84A1* switching unit can switch traffic to and from up to nine selective calling lines. A tenth line-termination circuit is provided as a spare. The number of stations on each line, and the total number of stations in a system, will vary greatly with the amount and kind of traffic. It was said before that the present average number of stations on a selective calling line is about ten. It is thought that more than 20 stations on a line will be found only rarely and then only when some of the stations are of a stand-by nature, with very light traffic requirements.

All transmission from the stations should be by automatic transmitters, from punched tape. While manual keyboard sending is possible on *83B2* lines, it prevents the realization of the benefits of an automatic system and is not recommended. TADS lines are not arranged for keyboard sending.

Trunks Between Switching Centers

One or more of the lines connecting to an *84A1* switching unit can be used as trunks to other switching centers. Way stations can be located along these trunks. However, because of the limited switching capacity of the *84A1* unit, the switching load offered by other lines must be reduced to allow for the heavier load introduced by the trunk. Before a multicenter system is projected, the probable traffic should be carefully analyzed to make sure that objectionable bottlenecks are not likely to occur.

The most successful operation of such a system would be expected when the local communities of interest around each switching center were large, while the demand for intercenter traffic was relatively small.

If more than two centers are used they must be in tandem; no provision for triangular trunking is made. Thus messages from one center which must be switched through a second to reach a third impose additional burdens on the switching capacity.

Functions of the Switching Unit

The block diagram of Fig. 3 illustrates the principal functions of the switching unit. For convenience in mapping the traffic flow across the page, the incoming lines are on the left and the outgoing lines on the right. It needs to be appreciated that an HDX circuit operates either incoming or outgoing (though not both at the same time) and consequently line no. *1* at both left and right ends of the diagram is the same line.

This means that, so far as the line itself is concerned, it cannot be sending into the switching center and receiving from it at the same time. On the other hand, on the office side of the reperforator-trans-

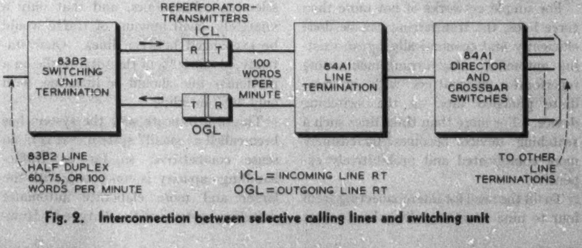

REPERFORATOR-TRANSMITTERS
ICL
R
T
83B2 SWITCHING UNIT TERMINATION
100 WORDS PER MINUTE
T
R
OGL
84A1 LINE TERMINATION
84A1 DIRECTOR AND CROSSBAR SWITCHES
83B2 LINE HALF DUPLEX 60, 75, OR 100 WORDS PER MINUTE
ICL = INCOMING LINE RT
OGL = OUTGOING LINE RT
TO OTHER LINE TERMINATIONS

Fig. 2. Interconnection between selective calling lines and switching unit

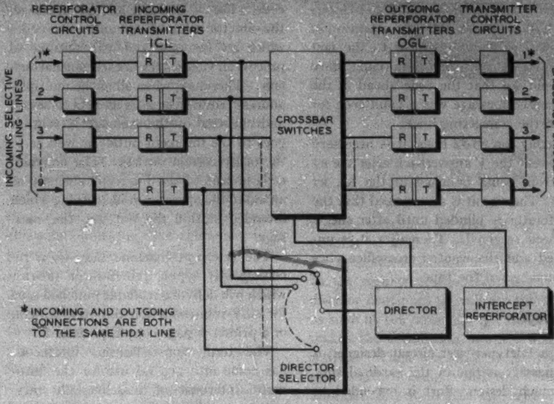

REPERFORATOR CONTROL CIRCUITS
INCOMING REPERFORATOR TRANSMITTERS
ICL
OUTGOING REPERFORATOR TRANSMITTERS
OGL
TRANSMITTER CONTROL CIRCUITS
INCOMING SELECTIVE CALLING LINES
OUTGOING SELECTIVE CALLING LINES
1*
2
3
9
R
T
CROSSBAR SWITCHES
DIRECTOR
INTERCEPT REPERFORATOR
DIRECTOR SELECTOR
*INCOMING AND OUTGOING CONNECTIONS ARE BOTH TO THE SAME HDX LINE

Fig. 3. Functions of the 84A1 switching unit

mitters, the transmission is essentially FDX. The ICL RT of line *1* can be sending a message crossoffice to one or more other line outlets at the same time that the OGL RT of line *1* is receiving a crossoffice message from some other incoming line.

For purposes of this description, it will be assumed that a message is being received on the incoming line RT from line no. *1* of Fig. 3. As soon as the first character of the incoming message is punched in the reperforator tape, the pivoted transmitter moves away from the punchblock and a loop of tape starts to form between them. A contact is operated which gives a "tape available" signal to the switching unit and makes a "bid" for the director. The director selector automatically searches until it finds line no. *1*, the "busy" incoming line. The director then causes the RT transmitter to step, one character at a time, as it reads the CDCs. When it has read a valid code, it proceeds to set up a crossoffice path to the indicated OGL RT and sends the CDC across the office. For multiple-address messages, the operation is repeated for each CDC until all the necessary crossoffice connections have been made. When a multiple-address message is switched, only the CDCs for a particular line appear in that line's outgoing reperforator.

After all of the CDCs comes the end-of-address code, which is carriage return, line feed, letters shift (CR LF LTRS). On this code the director is released, the incoming transmitter starts and the text of the message is sent across the office at 100 words per minute. It is received as punched tape in the reperforators of all of the OGL RTS which were connected in accordance with the addresses.

The OGL RT bid for a chance to send to its outgoing selective calling line is not made when the OGL pivoted transmitter first moves away from the punch as it receives a CDC from the director. The bid is withheld until the end-of-address code. This is necessary to prevent the outgoing lines from timing out in cases where the director encounters a delay in making a connection because one of the outlets it wants is already busy with another crossoffice connection.

When, however, after CR LF LTRS, the bids are made, the selective calling lines operate in their normal fashion, i.e., each transmitter waits until it is properly polled by the transmitter start circuit of the line, when it starts and sends the message to the line.

Finally the CDCs sent on the outgoing line select the one or more receiving stations to which the message is addressed, and the text of the message is printed at these stations only.

At the end of the text the disconnect code, figures shift, *H*, letters shift (FIGS H LTRS) disconnects the crossoffice paths in the *84A1* unit and prepares them to receive other requests for crossoffice connections. The disconnect code on the outgoing line disconnects the one or more receivers, put the stations in condition to respond to selective calling or transmitter start codes, and restarts the search pattern of the transmitter start circuit so that the next transmitter with a message to send may get its chance.

Trunk Route Codes

If the director reads a CDC for a station on a line out of a distant switching center, it will automatically supply an extra CDC, known as a trunk route code, which is required to advance the message toward its destination. The trunk route code is simply the CDC for the reperforator on the trunk at the distant center. This feature simplifies the format for the originating operator as it removes the necessity of determining in advance for each address which switching center serves its station.

If a message requiring a trunk route code is willfully intercepted, the trunk route code will be supplied in the intercept operation. This makes it possible to introduce the intercept tape at a way station on the trunk without further addition of CDCs.

Message Format

The successful performance of the various automatic operations involved in this travel of a message from an incoming line across the switching unit office to an outgoing line requires a considerable degree of rigidity in the format of the address and control codes. A sample of format is given in Fig. 4. A few words of explanation are needed.

As is described in the selective calling system papers,[1,2] all of the CDCs consist of 2-letter combinations. For timing and other reasons each CDC is followed by a LTRS character. Six of the letters of the alphabet are not used in CDCs, either because they are reserved for control codes, such as *V*, *M*, and *H*, or because they can be generated by a short open on the line (of one or more code element lengths), such as *T* and *O*, and hence are more likely to be falsely generated by hits than characters with more complicated combinations of marks and spaces.

The letter *Y* is reserved for the second letter of all codes used to connect the

FORMAT IN THE ORIGINAL TAPE

10 LTRS LE LTRS BY LTRS AD LTRS CR LF LTRS
(Message Text) FIGS H 7 LTRS

FORMAT IN THE OGL RT TAPE ON THE LINE WITH STATION AD

AD LTRS CR LF LTRS (Text) FIGS H LTRS

Fig. 4. Typical format for an interline message from station LE to station AD (see Fig. 1)

reperforator sides of the switching unit RTs. This restriction allows a distinction between intraline and interline CDCs (to be described later), to be made simply and cheaply.

To summarize, six of the letters of the alphabet, *T*, *O*, *M*, *V*, *H*, and *Y*, cannot be used in CDCs. If any one of these letters appears at the director in a CDC position, that address is interpreted as nonvalid and goes to intercept.

This leaves 20 letters available for CDCs. Consequently, subject to some further occasional and minor restrictions on the selective calling lines, a total of 400 different 2-letter codes is available for the system.

Referring now to Fig. 4, the multiple LTRS characters at the beginning and end are an aid to tape handling at the sending station transmitter. (Surplus LTRS characters at the end of a message are promptly discarded by the director.) The code LE cuts on the sending station so that a home copy is obtained while the message is being sent on the line. For an *83B2* line this is optional; on a TADS line the sending station CDC must be supplied. The code BY connects the *84A1* reperforator. The format in the OGL RT tape is shown in the second example of Fig. 4. The station codes are those used in the typical layout of Fig. 1 where the geographical relations may be seen.

On *83B2* lines, a further qualification is needed because of the *V* answerback used for circuit assurance. On all codes valid for the line, which include *LE* and *BY* in this illustration, the sending transmitter (at station *LE*) stops and waits for the return of a *V* from the called station. Such *V*s are returned from stations *LE* and *BY*. But *AD* is nonvalid for the sending line so that there will be no answerback after it. Here use is made of the letter *Y* in the RT cut-on code. This *Y*, combined with the following *V* (the answerback), operates a contact at every station on the line, including the particular station which happens to be sending, so that thereafter the transmitter will not stop, and no answerback is awaited. Thus the codes, nonvalid for the sending line but needed for the interline connections at the other side of the office, can be punched in the RT tape without unwanted system operations.

It follows that all intraline codes should be punched in the tape before the RT cut-on code. If one of them were added inadvertently after the RT code it would go on into the switching unit and appear at the director. Here, because of an "exclusion" feature which prevents a code from being sent back on the line from which it came, it would go to intercept. This would call the attention of the switching unit operator to the fact that a format error had been made, since he could see that the code ahead of the intercepted message was a valid code for one of the connecting lines.

Further, on *83B2* lines, it is necessary to prevent the *V* answerback after the RT CDC from being punched in the ICL RT tape. The circuit is so arranged that the reperforator is blinded until after the *V* has been received. Thereafter it is unblinded and the wanted crossoffice CDCs are punched in the tape.

Intercept

To a teletypewriter circuit designer, a lost message is one of the cardinal sins, and much design effort is expended to devise means to prevent this happening. One way is to provide an automatic means for intercepting messages which cannot be delivered because of a wrong or garbled address or because the receiving station is in trouble or out of service.

The *84A1* system is well provided with intercept features. In the first place, each connecting line has its own line intercept arrangement. Undeliverable messages addressed from one station to another on the same line are automatically intercepted. Incidentally, this feature makes possible the unattended operation of outlying stations with reduced chance of message delays resulting from undeliverable messages.

Second, the switching unit itself will intercept any messages with a CDC which the director does not recognize as a valid code. For a multiple-address message, the director will intercept only those codes which are nonvalid and will continue to process all valid codes to their normal outlets. Therefore, when all of the multiple-address crossoffice paths have been set up and the test of the message is sent, it goes to the intercept outlet as well as to the other correct outlets. The nonvalid code is sent to the intercept machine as an aid to the operator in identifying which intended station did not get the message.

The intercept machine may be a receiving-only typing reperforator (ROTR), which will deliver a readable punched tape for easy reintroduction into the system, or a printer if page copy is preferred.

The term "miscellaneous intercept" has come into general use for the automatic intercept of undeliverable messages. Another type of intercept, known as "willful intercept," takes place when an operator performs a manual operation to intercept deliberately messages for particular destinations. This might be used, for example, when an operator knows that a certain station is to be out of service for maintenance work.

The *84A1* switching unit has a key for each station on its satellite lines which when operated by an attendant diverts to intercept all crossoffice messages directed to that station. In addition there is a group of line keys, each of which will divert the traffic for all of the stations on a particular line.

Another provision for willful intercept is made at the incoming RT cabinet. By the operation of a key, designated FIGS H & INTCP, in conjunction with a BUSY key,

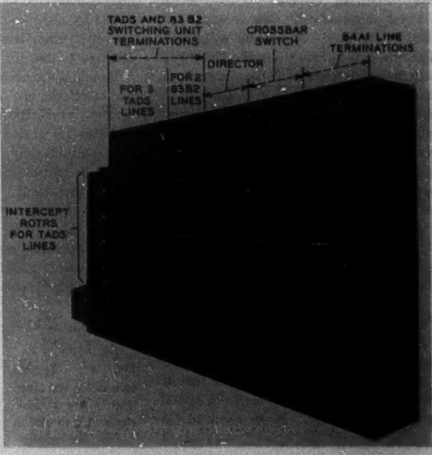
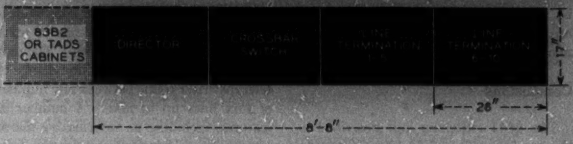


Fig. 5. The 84A1 switching unit with 83B2 and TADS line equipment



CABINET HEIGHT: 7 FEET

Fig. 6. Cabinet lineup for 84A1 switching unit

an operator may divert to intercept a message which is waiting in the ICL RT. This may be used to relieve a director tie-up caused by a busy OGL RT.

Director Operation on Nonvalid Letters and Function Characters

The director operations are performed in two fan circuits, called the first fan and the second fan. The normal operation of the director on CDCs (two valid characters plus one LTRS character) is to process the first character in the first fan, the second character in the second fan, and then to send through the LTRS character. In the case of nonvalid letters or function characters several different things may happen, depending on the circumstances.

If a LTRS character appears in the first fan it will be discarded. This is to dispose of the LTRS characters which may precede the CDCs and to assure that after these LTRS the CDCs will be in step, i.e., first CDC character to the first fan and second to the second fan. If a LTRS character should appear in the second fan, because of a malfunction which threw the CDCs out of step, the message would go to intercept.

Any of the following characters in either the first or the second fan are nonvalid as CDC characters and so will send the message to intercept: Blank *T*, *O*, *M*, *V*, FIGS, and Space.

Either CR or LF appearing in the first fan is recognized as end-of-address. The director is retired and does not process the following characters as CDCs. If CR or LF appeared first in the second fan, not having appeared in the first fan, the message would go to intercept.

These features increase the degree of protection provided in the switching unit against lost messages.

Alarms, Keys, and Lamps

The control panels in the *84A1* equipment cabinets have a number of alarm lamps and control keys designed so that an attendant may discover quickly the source of trouble and take steps to correct it. A common alarm lamp and bell, interlocked with the individual alarms, alerts the attendant in case of trouble anywhere in the switching center.

As an illustration, the excess-wait alarm in the director cabinet serves notice that the director has been waiting more than 2 minutes for a crossoffice connection to an outlet which it cannot reach because the outlet is busy. A lamp shows which outlet is causing the delay. There are two possible outcomes in this situation, depending on what the attendant does with the AP (automatic push to intercept) key. In the regular position of this key, the director continues to wait for the outlet; as soon as the outlet is free, the director makes the connection and the alarm is retired. When the key is in the AP position, the message which was waiting is automatically routed to intercept at the end of the 2-minute wait period, and the director is released to process the next incoming message. The excess-wait alarm does not appear in this case, but the intercept alarm tells the attendant that some action on the intercepted message is required.

A key is provided to stop the action of the director entirely, which may be needed to patch in the spare director or during maintenance operations.

A number of the alarms and controls are associated with reperforator-transmitters. Basic are the low-tape alarms, warning the attendant that the tape supply needs to be replenished. Another alarm warns if the tapes are torn or tangled. HOLD and BUSY keys are provided for the incoming and outgoing RTS to hold back messages during maintenance operations on any of the machines. If these keys are operated in the middle of a message, in most cases the action does not take place until the current message is completely sent or received. However, immediate stop can be had if needed.

Supplementing the switching unit alarms are the alarms normally provided with the conecting *83B2* and TADS lines. With the combination, there cannot be a major delay in system operations without a warning to the attendant.

Testing

A test cabinet and test bench for checking the circuits and for testing the machines, are offered as optional equipment with an *84A1* switching unit. The test cabinet contains line and equipment jacks for access to the line circuits, and trunks to the *84A1* switching unit equipment. The test position may be equipped with a monitoring printer, a transmitter-distributor for sending test messages from punched tape, and transmission measuring sets for examining the quality of signals.

The *84A1* switching unit cabinets, in addition to the trunks to the test cabinet, also have test panels behind the front covers normally available for use only by maintenance personnel. These contain various lamps and keys for use in testing the various switching unit components.

Spare Equipment

A spare director is provided in the director cabinet as insurance against lost time in case of trouble and to allow testing and maintenance work without service interruption. Convenient patch cords and connectors inside the cabinet permit easy and fast patching.

One spare line termination circuit is provided in the first line cabinet. It is expected that each switching unit would have some spare reperforator-transmitters to allow for maintenance work.

Equipment Arrangements

The control circuits of the *84A1* switching unit are mounted in four steel cabinets, three of which are shown in Fig. 5. The floor plan is shown in Fig. 6. For fast and easy installation, all connections between the *84A1* cabinets are made by plug-in connectors. Because of the lengths of the cords provided, the cabinets must be arranged in the order shown in Fig. 6.

Matching cabinets, which may be mounted in the same line as the *84A1* cabinets if space permits, contain the *83B2* and TADS switching unit terminations. Three of these are shown in Fig. 5.

The *14*-type reperforator-transmitters are mounted in the standard cabinets normally supplied for them. The ICL and OGL RTS for one line are both mounted in the same cabinet, with the incoming line connected to the upper level and the outgoing to the lower.

The test cabinet and test bench may

be in the same room or, if desired, in a separate room devoted to testing and maintenance.

Conclusions

The *84A1* switching unit described in this paper was designed to be a relatively inexpensive automatic teletypewriter message switching unit for businesses which need more automatic communications than can be given by one or more separate selective calling lines. The unit provides for the automatic interchange of teletypewriter traffic between any number of HDX lines up to nine. A number of features, including an informative alarm system, have been designed for the fast, smooth flow of traffic with a minimum of service interruptions.

References

1. AUTOMATIC DISPATCH SYSTEM FOR HALF-DUPLEX TELETYPEWRITER LINES, J. L. Maxwell, H. K. Farrar. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, Sept. 1956, pp. 441–45.

2. THE 83B1 TELETYPEWRITER SELECTIVE CALLING SYSTEM, C. W. Smith, C. J. Votaw, A. L. Whitman. *Ibid.*, May 1959, pp. 119–24.

3. A FULL AUTOMATIC PRIVATE-LINE TELETYPEWRITER SWITCHING SYSTEM, W. M. Bacon, G. A. Locke. *Ibid.*, vol. 70, pt. I, 1951, pp. 473–80.

Designs for Large Organic Reactor Power Plants

G. H. BOSWORTH
NONMEMBER AIEE

W. E. PARKINS
NONMEMBER AIEE

ONE of the types of nuclear reactors now under consideration for use in large central station plants makes use of an organic fluid as moderator and coolant. The organic fluid technology as applied to a nuclear reactor has been under development for several years. In addition, the Organic Moderated Reactor Experiment (OMRE) has been constructed and operated at the Nuclear Reactor Testing Station in Idaho. A number of preliminary designs have been prepared for organic reactor plants for the production of electric power as well as for the propulsion of merchant ships and the production of process steam for industry.

Perhaps the most suitable application, and certainly the one receiving the greatest present attention, is the use of organic reactors in central station power plants. The Atomic Energy Commission (AEC) recently approved a contract for the final design and construction of a small organic plant for the production of 11.4 net emw (electrical megawatts) for the city of Piqua, Ohio. This plant for Piqua, the OMRE, and the preliminary design for a 150-emw plant have been previously described.[1]

Recently the AEC sponsored a study by the Bechtel Corporation and Atomics International for a large, optimized organic-cooled reactor plant. The study was carried out simultaneously with similar studies by other organizations on the pressurized-water, boiling-water, and heavy-water-type reactor plant.

On the organic reactor project, an engineering survey was first conducted to determine which type of moderator showed the greatest promise for low cost power when used in a reactor cooled by an organic fluid. The survey, which is discussed in a subsequent section of this paper, indicated that a reactor moderated by the same organic fluid used as the coolant, possessed the greatest potential. An effort was then devoted to the preliminary design of a 300-emw-output plant with a secondary effort on the design of a similar plant for 75 emw. Finally, the preliminary design of a completely conventional 305-emw power station was prepared for a comparison of performance based upon identical site conditions. The results of these studies, together with a summary report on the work involving nuclear reactors, have been described in reports[2–6] now made available by the AEC. In this paper, some of the pertinent background technology for organic reactor systems is given, together with brief descriptions of the plant designs prepared and principal conclusions reached during the recent organic-cooled power reactor study.

Background Technology

Of the various organic fluids which have been considered for use in nuclear reactors, greatest attention has been devoted to commercial mixtures of the isomers of terphenyl. This material has been shown to be relatively stable at high temperatures and in the presence of radiation. It is available in large quantities at a cost of less than 20 cents per pound and is made by a process which uses benzene as the raw material. Most of the reactor design work has been based upon use of a terphenyl mixture known as Santowax-R, which boils at a temperature slightly above 700 F (degrees Fahrenheit) and melts over a range of temperatures from about 240 to 300 F.

The principal advantages of organics such as Santowax-R, when used in a nuclear reactor, are the relatively low vapor pressure and the negligible corrosive effect on nuclear fuels and on standard materials of construction. The principal disadvantages are relatively poor heat transfer properties and the gradual chemical decomposition which occurs in the presence of radiation. This decomposition requires purification of the coolant and the addition of fresh make-up material.

HEAT TRANSFER

Considerable experimental work has been carried out on many aspects of organic fluids, including the heat transfer rates attainable and the rates of radiation decomposition under practical reactor operating conditions. One of the important questions on the subject of heat transfer has been relative to changes in the heat transfer coefficient which might occur over long periods of time due to build up of any surface film from the organic fluid. Measurements in heat transfer loops and in the OMRE have demonstrated that relatively high temperatures, for example 850 F, may be employed at the heat transfer surface without significant long-time changes in the coefficient. The relatively low value of the coefficient itself, however, usually dictates the use of a high-temperature drive, perhaps 150 F, across the heat transfer film. Experiments have shown that in a practical reactor design the bulk organic can exit from the reactor at a

---

Paper 59-1132, recommended by the AIEE Nucleonics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted June 15, 1959; made available for printing July 28, 1959.

G. H. BOSWORTH is with the Bechtel Corporation, San Francisco, Calif.; W. E. PARKINS is with Atomics International, Canoga Park, Calif.

temperature in excess of 600 F, and that steam can be produced under superheat conditions in heat exchangers at temperatures of 600 F or somewhat above.

One of the heat transfer conditions which is being explored as a result of the recent organic-cooled power reactor study has been the use of subcooled nucleate boiling. Under this condition the heat transfer surface temperature operates above the boiling point of the fluid for the existing pressure, while the bulk fluid is at a temperature below the boiling point. Vapor is produced by boiling at the surface and it subsequently condenses in the bulk fluid. The advantage of this heat transfer mechanism is that good heat transfer rates can be achieved at low fluid velocities as compared with the velocities which would be required under nonboiling, or forced convection, conditions. This has the effect of reducing pressure drops in the organic circulating system and of lowering pumping power requirements. Experiments are presently in progress to establish the limiting conditions under which nucleate boiling heat transfer can be applied in a reactor core.

ORGANIC DECOMPOSITION

The operating conditions of current interest in organic-cooled power reactors are well below the temperature range where thermal, or purely pyrolytic, decomposition is appreciable. This pyrolytic decomposition is primarily related to the temperature of the bulk fluid, whereas the limiting conditions are presently being established by the temperature of the heat transfer surface.

The other type of decomposition, that which is due to the effects of radiation, is still present in any operating organic reactor. This decomposition results in the formation of volatile light hydrocarbons as well as in polymers, or high boilers, of lower vapor pressure than the organic feed material. The principal contributors to this radiation decomposition in an organic reactor core are gamma rays and energetic neutrons.

The radiation decomposition rates of organic fluids have been carefully measured in the OMRE as well as in separate in-pile experiments. This rate has been shown to decrease as the equilibrium high boiler content of the circulating organic increases. For this reason, as well as to reduce the rate of purification requirements in any practical organic reactor it is planned to operate with a high boiler content, in the neighborhood of 30%. Under such conditions the make-up requirement for the addition of fresh organic to the system amounts to approximately one pound per thermal megawatt-hour of operation. The cost contribution of this operational requirement amounts to approximately 0.6 mill/kwhr (per kilowatt-hour). It is expected that future reductions in the cost of the commercial organics and the eventual use of inhibitors in reducing radiation decomposition rates will appreciably lower this organic make-up expense.

REACTOR FUEL

Since the organic is a noncorrosive liquid, it is possible in principle to use any of the available nuclear fuels. It is, of course, desirable to clad the fuel material with some metal to prevent the escape of fission products into the coolant.

Previous designs of organic reactors have given principal emphasis to a fuel of uranium metal alloy clad with aluminum. An alloy consisting of uranium with 3.5% molybdenum and 0.1% aluminum has been recommended as one possessing good properties of stability for relatively high burn-up at the required operating temperatures. This material may be metallurgically bonded to aluminum cladding through a thin nickel-plated interface. To improve the heat transfer in the organic fluid, most designs have been based on an extruded aluminum cladding incorporating short fins to increase the surface area. Such a surface is arranged so that the fins are parallel to the direction of fluid flow, and in a practical case the fin configuration more than doubles the heat flux attainable from the fuel surface for given temperature conditions.

More recently, again as a part of the organic-cooled power reactor study a recommendation was made for the use of uranium dioxide (UO_2) as the fuel material. In this case the lack of strength of UO_2 itself led to the choice of APM (aluminum powdered metal, made by sintering aluminum and a small percentage of aluminum oxide) as the cladding material. A fuel rod made from these materials consists of a column of cylindrical pellets of UO_2 sealed with a helium gas atmosphere inside of a long tube of APM. Again, the design incorporates extruded longitudinal fins on the APM tube. The complete fuel element consists of a large number of the APM-clad UO_2 rods assembled into a box of square cross section through which the organic coolant flows.

The use of the APM-clad UO_2 is compatible with the high-temperature surface conditions desired. In addition, the UO_2 has been shown by a number of experiments to be a good stability fuel capable of achieving very high burn-up. The heat fluxes attainable from the UO_2 rod, as limited by the melting point of UO_2 at its center, are compatible with those which can be employed with the organic fluid.

Organic-Cooled Power Reactor Study

The study on a large organic-cooled power reactor plant by the Bechtel Corporation and Atomics International was initiated in December 1958. One of the guiding principles for this study was the objective of optimizing the size of the nuclear power station with regard to the anticipated growth of utility systems and to the production of minimum cost power. The plant was to be based on the most advanced technology available, consistent with a start of construction in July 1960.

REACTOR SELECTION

One of the first tasks was to select the moderator material leading to the design of highest over-all performance, when used in a large reactor cooled by an organic fluid. The moderating materials included in this evaluation were beryllium, beryllium oxide, zirconium hydride, heavy water, light water, graphite, and organic fluid. Survey nuclear and heat transfer calculations were conducted to determine the relative sizes for reactors utilizing these different moderators, and designs were prepared to assess the principal engineering problems.

Estimates of the differences in capital and operating and fuel costs were made to indicate differentials in the cost of power from the several systems examined. The reactor types were evaluated primarily from the point of view of incremental costs, although factors such as technical status, required development program, and feasibility from the standpoint of start of construction in July 1960 were necessarily considered.

Of the various concepts examined, the organic-moderated and cooled reactor system showed the lowest total power cost. In addition, this system resulted in the simplest design and required the least research and development effort. It was apparent that the organic-moderated type was also in a position to benefit the most from future anticipated reductions in coolant make-up cost. It was concluded, therefore, that the further design effort should be directed toward an advanced organic-moderated and cooled reactor plant, designated the OMCR.

The selection studies indicated the

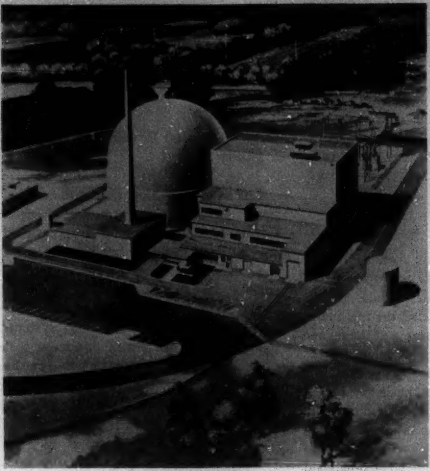
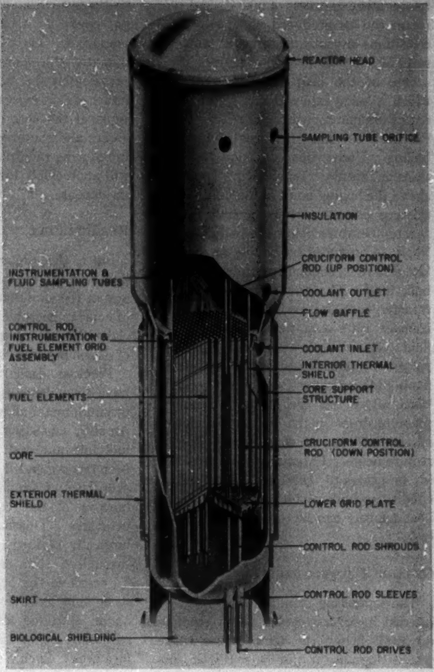


Fig. 1 (above). Perspective of 300-emw reactor plant

Fig. 2 (right). The 300-emw organic reactor


REACTOR HEAD
SAMPLING TUBE ORIFICE
INSULATION
CRUCIFORM CONTROL ROD (UP POSITION)
INSTRUMENTATION & FLUID SAMPLING TUBES
COOLANT OUTLET
FLOW BAFFLE
CONTROL ROD, INSTRUMENTATION & FUEL ELEMENT GRID ASSEMBLY
COOLANT INLET
INTERIOR THERMAL SHIELD
CORE SUPPORT STRUCTURE
FUEL ELEMENTS
CRUCIFORM CONTROL ROD (DOWN POSITION)
CORE
EXTERIOR THERMAL SHIELD
LOWER GRID PLATE
CONTROL ROD SHROUDS
SKIRT
CONTROL ROD SLEEVES
BIOLOGICAL SHIELDING
CONTROL ROD DRIVES

second choice of moderator for an organic-cooled power reactor to be graphite. Such a graphite-moderated organic-cooled power reactor plant could be constructed in a large size and would produce power at costs approaching those from a reactor plant using organic fluid as the moderator. An attractive aspect of the graphite-moderated design is that, because of the use of externally manifolded fuel channels, this system offers a very promising future means of producing nuclear superheat. For this purpose certain of the outer fuel channels would be cooled by the steam produced in steam generators heated by the organic circulating through the other fuel channels.

Unit Capacity Selection

A unit rating of 300 emw net was selected as the size best meeting the various criteria stipulated by the AEC for the reference plant. The primary considerations leading to this selection were

1. Extensive discussions with turbine manufacturers revealed that, for the optimum steam conditions, a 300-emw unit would be the maximum size which could be made available, based on present and near-future turbine design technology.
2. Trends in the utility industry continue in the direction of constructing larger units, in line with the growth of markets and load centers. Units of 300-emw rating will be increasingly attractive in the future and are in the size range expected to constitute a significant fraction of the total new plant capacity to be constructed.
3. A single organic reactor installation is capable of supplying the required steam flow at suitable conditions for a 300-emw unit. Actually, a single reactor is capable of supplying much larger plants without imposed engineering limitations.
4. The emphasis on minimum generation cost logically leads to selection of large units to take advantage of economies inherent in large pieces of equipment and to distribute construction and operating costs over the maximum output of electrical power. The maximum size of available turbine essentially sets the plant size.

300-Emw OMCR Plant

The 300-emw OMCR plant is illustrated in Fig. 1. The reactor building is a 190-foot-diameter self-supporting hemispherical dome enclosing the reactor, primary coolant loops, steam generators, and fuel storage facilities. This building provides a low-pressure containment barrier to prevent any uncontrolled release of radioactivity to the environment. The low reactor building wing in front of the dome houses reactor and primary loop auxiliaries and is surmounted by the ventilation exhaust stack. The large building to the right of the dome houses more conventional equipment associated with the power generating facilities. A central control room, offices, shop, warehouse, and personnel facilities are located in the turbine building.

Reactor

A cutaway view of the reactor for the 300-emw OMCR is shown in Fig. 2. The reactor vessel is unclad and is fabricated of low-alloy steel. It is a cylindrical vessel 67 feet, 7 inches, in height with a welded dished head at the bottom and a removable dished head at the top. The upper section of the tank has an inside diameter of 17 feet and a wall thickness of 2 inches. The lower section, which contains the core, is 13 feet, 6 inches in inside diameter and has a wall thickness of $1^{1}/_{2}$ inches. The vessel is designed for 100 psig (pounds per square inch gage) at 750 F. Sixty-one cylindrical control rod thimbles penetrate the lower head and extend into the subpile room where the control-rod drive mechanisms are located.

The reactor core contains fuel element assemblies of APM-clad UO_2 rods. There are 376 fuel-element positions on a square lattice in the core. Active fuel elements occupy 352 locations with the remaining 24 occupied by 22 dummy elements and and 2 neutron sources. The core is surrounded by a steel inner thermal shield, which also serves as a core-supporting structure. Steel grid plates are used below and above the core with the upper grid containing the necessary provisions for routing in-core instrumentation thermocouple and tubing lines.

The organic coolant enters the core tank at a temperature of 550 F through four 30-inch-diameter inlet nozzles. It flows downward through the annulus between the reactor vessel wall and the inner thermal shield, mixes in the lower plenum, flows upward through the fuel element boxes, and from the upper plenum out through four 30-inch nozzles. Flow distribution through the fuel elements is controlled by individual orifices located at the lower end of each element. These orifices are sized to obtain a flow distribution which will result in a uniform coolant outlet temperature of approximately 675 F from all elements when the reactor is operating at full power. Heat transfer throughout the central region of the core is by means of subcooled nucleate boiling.

The reactor vessel is located within a cavity formed by the concrete walls of the biological shield. A steel outer thermal shield immediately surrounds the reactor vessel in the vicinity of the core, and a commercial thermal insulation occupies the remaining space between the core tank and the concrete wall of the cavity.

Sixty-one cruciform-shaped control elements are used in the reactor core to regulate power, to assist in adjusting the radial and axial core power distribution, and to serve as safety rods for any emergency shutdown. These elements are arranged in such a pattern that in the control region of the core there is one control element for every four fuel elements. Stainless steel is used as the cladding material for the cruciform sections, with the neutron-absorbing materials being Eu_2O_3 in some control elements and boron steel in others. The elements are actuated from the subpile room below the reactor vessel and are removed or replaced from above the core.

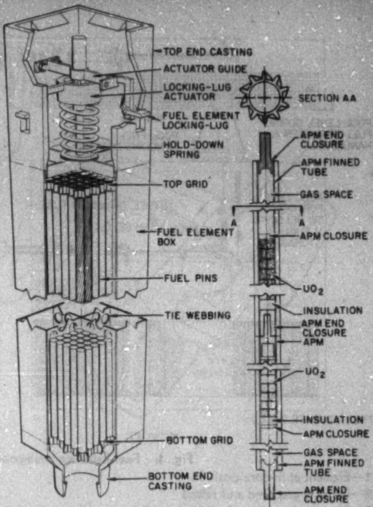


Fig. 3. Fuel element: 100 0.3-inch-diameter UO_2 fuel pins clad with APM in a stainless-steel box

Fuel

The UO_2 fuel is enriched to 2.5% U-235 and is clad with finned APM tubing. Each fuel element, as illustrated in Fig. 3, consists of 100 vertical fuel rods assembled in a long 0.035-inch-thick stainless-steel box measuring $5^1/_8$ inches on a side. The over-all length of the fuel-element assembly is approximately 15 feet, while the length containing fuel within the assembly is approximately 12 feet. Each fuel rod is made of two sections 7 feet long, sealed by APM end plugs. The UO_2 pellets are 0.3 inch in diameter and 0.6 inch long. The fins on the APM cladding tube are fabricated to have a slight helical pitch to prevent meshing and to assure proper spacing when assembled within the fuel element box.

To achieve the most efficient fuel utilization it is necessary to obtain the maximum possible burn-up for each fuel element in the reactor. By properly programing the replacement of fuel elements, it is possible to obtain an average fuel burn-up of 15,000 thermal mw-days per metric ton of uranium. This is based upon a fuel loading schedule which makes use of three annular zones of fuel elements. During refueling periods the fuel is moved from one zone to another with the new fuel loaded initially in the peripheral zone. The high burn-up fuel in the central region of the core assists in uniformizing the radial power distribution.

To remove fuel from the reactor vessel the handling arrangement shown in Fig. 4 is used. The reactor must be shut down, cooled to a lower temperature, depressurized, and the top head removed. A mechanism is provided which engages the upper end of any given fuel element, raises it within the pool of organic above the core, and tilts it to a horizontal position where the fuel element is discharged into a tube leading to the water storage pool. During the transfer to the water storage, separation between the two fluids is appropriately maintained by valves, and the element is cleaned through the action of a solvent wash. This under-liquid handling system has the advantage of maintaining the fuel element immersed, and hence adequately cooled, at all times during the transfer process.

Organic Coolant System

The coolant is specified as Santowax-R containing approximately 30% of high boiler compounds. This fluid is circulated in four parallel heat transfer loops located in separate compartments around the reactor vessel, see Figs. 5 and 6. All piping and components in the organic system are fabricated from low-carbon steel. Each loop is capable of being isolated from the remainder of the system by blocking valves. The pump in each loop is a vertical, 2-stage turbine-type unit rated at 31,200 gallons per minute and uses a 1,500-horsepower motor drive.

The organic liquid circulating in each loop gives up its heat in a U-shaped horizontal tubular steam generator and in a superheater of similar design located above the steam generator. The organic fluid flows in the shell side in the superheater, and in the tube side in the steam generator. At full capacity the four primary loops produce a total of 3,640,000 pounds of steam per hour at 600 psig and 650 F at the turbine throttle. For control of heat input to the steam generators at different power levels, each loop incorporates a valved bypass line around the steam generator and superheater combination.

Each of the four steam generators is provided with two power-operated relief valves, with each valve capable of removing approximately 120% of the steam-generator design steam flow. In addition, each steam generator is provided with safety valves as required by the American Society of Mechanical Engineers boiler code. The primary coolant loops are protected from any excess pressure which would result from rupture of a steam generator tube by a series of relief valves on the main coolant lines.

The plant is so designed that emergency decay heat removal from the core can be accomplished by natural convection circulation through any two of the four steam generators. To avoid dependence on availability of power during this period sufficient feedwater is delivered to the generators by small turbine-driven emergency feed pumps, taking suction from the condensate storage tank and utilizing a part of the steam generated by the decay heat.

Auxiliary equipment for the organic system provides for purification and degasification of the organic and means for disposing of the gases produced. The degasification system removes any water vapor, hydrogen, and light hydrocarbons in the organic by passing a small side stream of the fluid through a low-pressure tank. The purification system separates the reusable coolant from the high boiler material by passing a normally continuous side stream through a single stage of fractionation. New coolant make-up material is added at a rate to compensate for the removal of the high boiler and gaseous material. This make-up rate during full power operation is approximately 1,105 pounds per hour. Gases removed by the degasification system and by the purification system are appropriately filtered, diluted by the building ventilation air stream, and exhausted through the stack. Special equipment is provided to remove fission products from the gases in the event of fuel-element cladding failures. Any radioactive material removed with the high boiler reject is allowed to decay to safe levels for waste disposal by burning, through the use of storage tanks sized for a maximum of 106 days of storage during full-power operation.

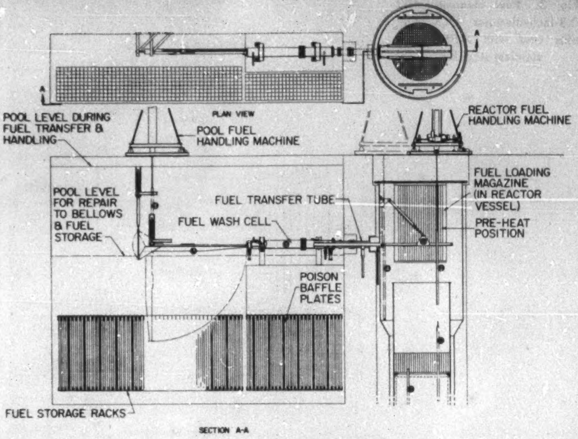

PLAN VIEW
POOL LEVEL DURING FUEL TRANSFER & HANDLING
POOL FUEL HANDLING MACHINE
REACTOR FUEL HANDLING MACHINE
FUEL LOADING MAGAZINE (IN REACTOR VESSEL)
POOL LEVEL FOR REPAIR TO BELLOWS & FUEL STORAGE
FUEL TRANSFER TUBE
FUEL WASH CELL
PRE-HEAT POSITION
POISON BAFFLE PLATES
FUEL STORAGE RACKS
SECTION A-A

Fig. 4. Fuel-handling arrangement

1—Element at in-core position
2—Element grappled and raised
3—Element in position for travel to pivot assembly
4—Pivot assembly in vertical position, element in place
5—Pivot assembly in horizontal position, pusher mechanism places element in wash cell
6—Element placed in wash cell
7—Pool pivot assembly, element withdrawn and placed in vertical position by pivot assembly
8—Fuel element lifted from pivot assembly and placed in storage rack

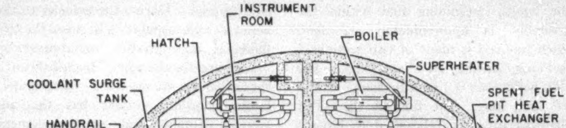

INSTRUMENT ROOM
HATCH
BOILER
SUPERHEATER
COOLANT SURGE TANK
SPENT FUEL PIT HEAT EXCHANGER
HANDRAIL
PUMPS
PIPEWAY
COOLANT PUMP
SPENT FUEL PIT
PIPEWAY
TURBINE BUILDING
TURBO-GENERATOR
ELEVATOR
REACTOR
SPENT FUEL PIT CENTRIFUGE
0' 5' 10' 15' 20'

Fig. 5. Plan arrangement of primary loops

Vol. Pt. 1 1959

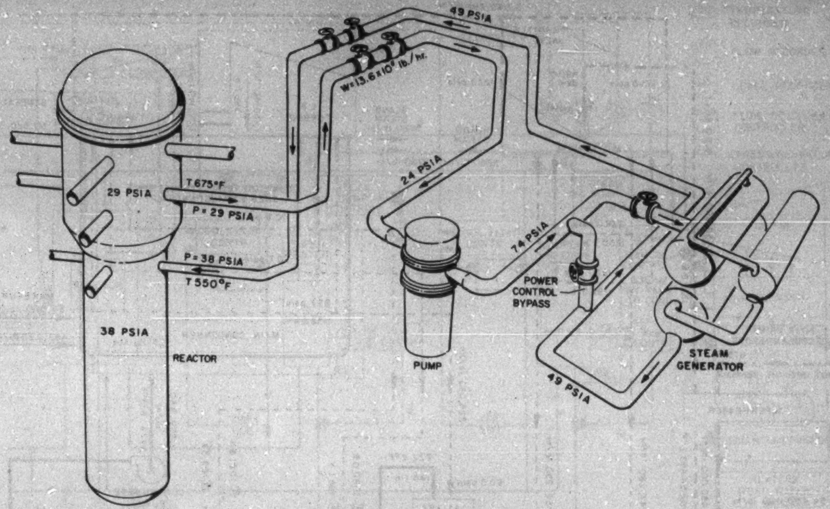


Fig. 6. Single primary loop schematic

TURBINE-GENERATOR AND ASSOCIATED EQUIPMENT

A flow diagram and heat balance for the steam system is given in Fig. 7. Nonradioactive superheated steam at 600 psig and 650 F drives a 300-emw (nameplate rating at 3.5 inches Hg [mercury] and 3% make-up) tandem-compound double-flow turbine-generator unit with a guaranteed throttle flow of 3,450,000 pounds per hour at 1,800 rpm. The gross and net outputs of the turbine-generator plant with 57 F river water and a condenser pressure of $1^1/_2$ inch Hg absolute are approximately 313 emw and 300 emw respectively. The gross heat rate is 10,414 Btu/kwhr. Exhaust steam from the turbine flows into a single-pass divided water box condenser located below the turbine. Screened river water is used to condense the steam which is reused in the system. The condensate is pumped through two low-pressure extraction steam heaters and a deaerating heater to increase its temperature. Boiler feed pumps taking suction from the deaerator force the feedwater through two high-pressure feedwater heaters and into the steam generators at a pressure of 635 psig and 400 F where it is again converted to steam, thus completing the cycle. Make-up requirements are met by evaporating treated river water in the auxiliary heating boilers and introducing the distillate into the deaerator.

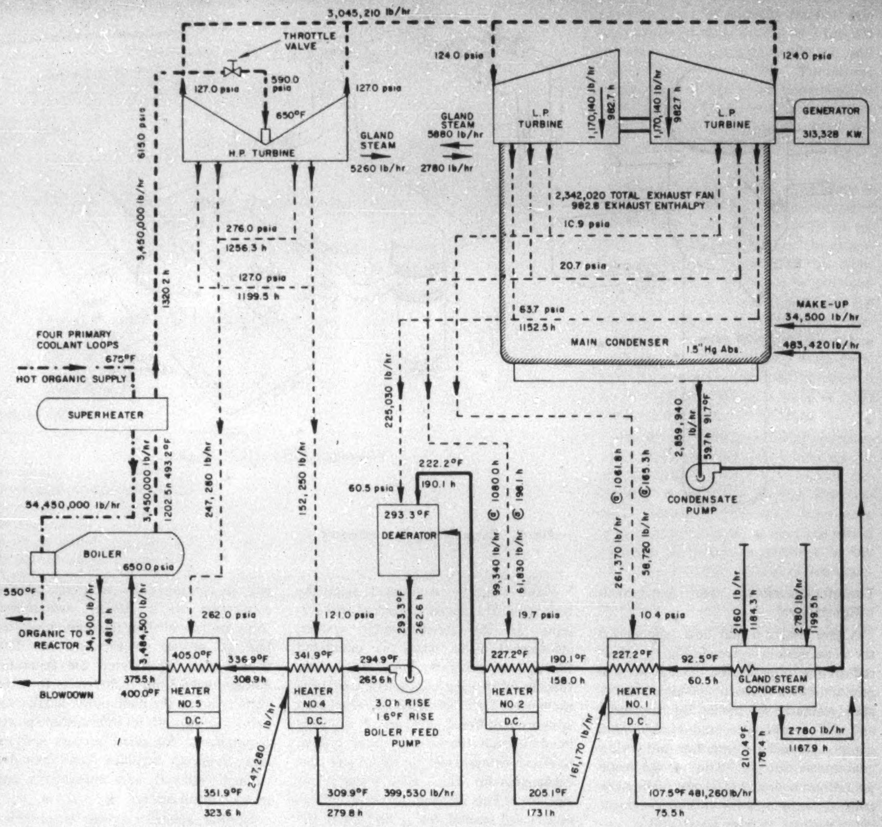
Plant auxiliary equipment requiring heat removal is served by one of two systems. In the service water system, chlorinated river water is circulated through the generator hydrogen coolers, turbine lubricating oil coolers, and heat exchangers for other separate closed-loop systems. Buffered condensate is circulated through the cooling water system to the air compressors, air dryer, and heat exchangers for which river water is not suitable. This cooling water circulates in a closed system, losing its heat to the river through a heat exchanger in the service water system.

The generator is of the totally enclosed, nonsalient pole type, is rated at 384,000 kva, 0.85 power factor, 326,400 kw, 0.64 short-circuit ratio, and generates at 16,500 volts, Y-connected. The stator is a double-winding type with the windings terminating in 12 terminal bushings, 6 on each end of the generator at the bottom. The generator is hydrogen-cooled at 45 psig. Two motor-driven full-capacity (one spare) exciters are provided, rated 1,800 kw at 500 volts d-c. A motor-operated exciter field rheostat and voltage regulator of the magnetic amplifier type are provided.

The generator is connected through metal-enclosed bus to the terminals of the low-voltage delta winding of the main power transformer. The high-voltage windings of the main power transformer will be connected by an overhead line connection to a 230-kv switchyard. Taps are provided on the main generator bus for service to the two auxiliary transformers which step the generator voltage down from 16,500 volts to 4,160 volts to serve the main power auxiliaries. Service at 480 volts for intermediate power requirements for plant process systems and building facilities are provided through 480-volt unit substations and motor control centers.

Normal auxiliary power is provided from the generator bus through two 1/2-capacity transformers which step the voltage down to 4,160 volts to serve main power auxiliaries. A start-up transformer delivers outside power at the same voltage and can carry up to half the unit load during a generator outage.

A highly reliable emergency power system is provided as required for essential loads to protect equipment, maintain ability to restore generation quickly, or ultimately to permit safe shutdown of the unit. Two diesel generators are available to back up power on the 480-volt emergency bus which also feeds a 125-volt battery system through two battery-charging inverter-type motor-generating sets. Certain critical auxiliaries which must be maintained to insure a programed reactor shutdown on loss of load can be supplied by these motor-generator sets during the interval required for one


3,045,210 lb/hr
THROTTLE VALVE
124.0 psia
124.0 psia
127.0 psia
590.0 psia
127.0 psia
650°F
H.P. TURBINE
6150 psia
3,450,000 lb/hr
1320.2 h
GLAND STEAM
5260 lb/hr
GLAND STEAM
5880 lb/hr
2780 lb/hr
L.P. TURBINE
1,170,140 lb/hr
982.7 h
1,170,140 lb/hr
982.7 h
L.P. TURBINE
GENERATOR
313,328 KW.
276.0 psia
1256.3 h
127.0 psia
1199.5 h
2,342,020 TOTAL EXHAUST FAN
982.8 EXHAUST ENTHALPY
1C.9 psia
20.7 psia
63.7 psia
1152.5 h
MAIN CONDENSER
1.5" Hg Abs.
MAKE-UP
34,500 lb/hr
483,420 lb/hr
FOUR PRIMARY COOLANT LOOPS
675°F
HOT ORGANIC SUPPLY
SUPERHEATER
225,030 lb/hr
2,859,940 lb/hr
59.7h 91.7°F
3,450,000 lb/hr
1202.5h 493.2°F
247, 280 lb/hr
152, 250 lb/hr
222.2°F
190.1 h
60.5 psia
293.3°F
DEAERATOR
99,340 lb/hr @ 1080.0 h
61,830 lb/hr @ 198.1 h
261,370 lb/hr @ 1061.8 h
58,720 lb/hr @ 165.3 h
CONDENSATE PUMP
54,450,000 lb/hr
BOILER
650.0 psia
550°F
ORGANIC TO REACTOR
34,500 lb/hr
481.8 h
3,484,500 lb/hr
BLOWDOWN
262.0 psia
121.0 psia
293.3°F
262.6 h
19.7 psia
10.4 psia
2160 lb/hr
1184.3 h
2780 lb/hr
1199.5 h
405.0°F
HEATER NO. 5
D.C.
375.5 h
400.0°F
336.9°F
308.9 h
341.9°F
HEATER NO. 4
D.C.
294.9°F
265.6 h
3.0 h RISE
1.6°F RISE
BOILER FEED PUMP
227.2°F
HEATER NO. 2
D.C.
190.1°F
158.0 h
227.2°F
HEATER NO. 1
D.C.
92.5°F
60.5 h
GLAND STEAM CONDENSER
2780 lb/hr
1167.9 h
210.4°F
178.4 h
351.9°F
323.6 h
247,280 lb/hr
309.9°F
279.8 h
399,530 lb/hr
205.1°F
173.1 h
161,170 lb/hr
107.5°F 481,260 lb/hr
75.5 h

$$\text{HEAT RATE} = \frac{3{,}450{,}000 \times 1320.2 - 3{,}484{,}500 \times 375.5 + 34{,}500 \times 481.8}{313{,}328} = 10{,}414 \text{ Btu/KWHR}$$

Fig. 7. Heat balance diagram

of the diesel generators to come on the line.

Control and Instrumentation

The reactor control and instrumentation system provides for automatic load following control in the power range and nonload following control during start-up and normal shutdown. In addition, it functions as the plant protective system providing automatic action to minimize or prevent damage as a result of misoperation or malfunction of equipment. There are instrumentation systems which furnish information on plant functioning and on the radiation levels in various areas. A diagram of the reactor plant control system is shown in Fig 8.

To provide the automatic load following control, system frequency regulates the throttle valve and voltage controls excitation. Feedwater control is assured by a conventional 3-element controller. Steam pressure ahead of the turbine throttle is used to control the flow of coolant by-passing the steam generators, thereby setting the rate of steam generation. The bulk temperature of the coolant from the reactor is used to control reactor power by automatic positioning of control elements. Up to 20% of the steam flow to the turbine can be manually bypassed to the condenser to facilitate start-up.

Emergency shutdown (scram) of the reactor is brought about by gravity insertion of all of the control elements into the reactor core. Scram, or power setback, wherein the control elements are driven at a controlled rate into the core, is automatically initiated by a number of signals. These are high or low coolant system pressure, high neutron flux, short neutron period in the start-up power range, high coolant outlet tempera-

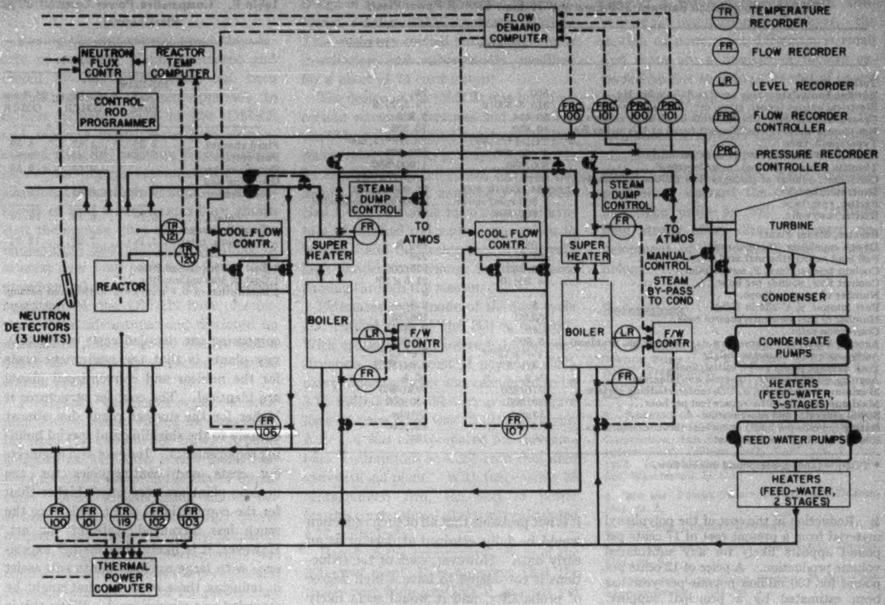

FLOW DEMAND COMPUTER
NEUTRON FLUX CONTR
REACTOR TEMP COMPUTER
CONTROL ROD PROGRAMMER
TR 121
TR 120
REACTOR
NEUTRON DETECTORS (3 UNITS)
COOL.FLOW CONTR.
SUPER HEATER
STEAM DUMP CONTROL
TO ATMOS
FR
BOILER
LR
F/W CONTR
FR
FR 106
FR 100
FR 101
TR 119
FR 102
FR 103
THERMAL POWER COMPUTER
FRC 100
FRC 101
PRC 100
PRC 101
COOL. FLOW CONTR.
STEAM DUMP CONTROL
TO ATMOS
FR
SUPER HEATER
BOILER
LR
F/W CONTR
FR
FR 107
MANUAL CONTROL
STEAM BY-PASS TO COND
TURBINE
CONDENSER
CONDENSATE PUMPS
HEATERS (FEED-WATER, 3-STAGES)
FEED WATER PUMPS
HEATERS (FEED-WATER, 2 STAGES)
TR TEMPERATURE RECORDER
FR FLOW RECORDER
LR LEVEL RECORDER
FRC FLOW RECORDER CONTROLLER
PRC PRESSURE RECORDER CONTROLLER

Fig. 8. Control diagram

ture, low coolant flow, low feedwater level, and loss of pumping power in two or more loops. In the event of a reactor scram, all of the bypass valves are trip-closed and the throttle valve tripped to an open position. This action minimizes thermal shock to the steam generator and automatically sets flow in the organic loops to the proper condition for decay heat removal by natural convection.

75-Emw OMCR Plant

The 75-emw OMCR was designed as an intermediate-size version of the larger plant just described. The 75-emw plant is based upon the same design criteria and incorporates essentially all of the same features as the 300-emw plant. Table I compares the specifications of the two plants.

There are two major design differences resulting from reduced size which merit mention. The 75-emw plant requires a much smaller primary system coolant flow; therefore, three loops of 24-inch-diameter pipe are used instead of four loops of 30-inch-diameter pipe as employed in the larger plant. Because the core height is less in the 75-emw reactor, the change in coolant temperature in passing through the reactor is 75 F instead of 125 F, where the outlet temperature has been maintained the same. This invites the use of 1,000-psig steam cycle, which has a lower steam rate, 9,822 Btu/kwhr, instead of 10,414 Btu/kwhr for the 300-emw plant.

Economics

The capital cost of engineering and construction of the entire plant for the first 300-emw unit has been estimated at $69,510,000, which is equivalent to $232 per net electrical kilowatt (ekw). Included in this total is the $300,000 which would be saved if no provisions were included for future plant expansion. These costs include escalation and interest during construction. For purposes of comparison with the reactor plant costs, as usually stated, the present-day cost excluding escalation and interest during construction is $184 per net ekw.

Several factors contribute to the relatively low cost of an organic reactor power plant. These are primarily related to the low vapor pressure of the coolant and the fact that it is noncorrosive. As a result, the reactor and its coolant system are designed for low operating pressures and are fabricated from low-carbon steel and other conventional materials. The use of the organic fluid as both the moderator and coolant adds to the simplicity of the system. In addition, the fact that the reactor is basically hydrogen moderated assures a compact core with resulting minimum costs for structure and shielding.

The estimated fuel cycle cost, at 80% plant capacity factor, varies from 2.67 mills/net kwhr (24.2 cents/million Btu) initially, to 2.45 mills/net kwhr (22.2 cents/million Btu) for the steady-state operation, based upon plant operation in the year 1963. Using an annual charge of 14% against the total plant investment, and adding the steady state fuel and other annual charges, the average cost of producing electric energy is estimated to be approximately 8.12 mills/kwhr at a plant capacitp factor of 80%.

Aside from further improvements in the design which may result in cost reductions, there are several areas in which substantial reductions in over-all power costs from an organic moderated and cooled reactor may be anticipated. These areas of potential improvement and their effect on the over-all cost for the 300-emw OMCR are as follows:

Table I. Comparison Between 300-Emw and 75-Emw OMCR Power Plants

	300 Emw	75 Emw
Turbine Plant:		
Turbine nameplate rating emw at $3\frac{1}{2}$ inches Hg	300	75
Electrical output (emw), gross/net, at $1\frac{1}{2}$ inches Hg	313.3/300.0	79.5/75.3
Gross plant heat rate (Btu/kilowatt-hour) at $1\frac{1}{2}$ inches Hg	10,414	9,822
Net plant heat rate (Btu/kilowatt-hour) at $1\frac{1}{2}$ inches Hg	10,832	10,309
Type/speed, rpm	TCDF*/1,800	TCDF*/3,600
Guaranteed throttle flow, pounds per hr	3.45×10^6	882,000
Throttle steam conditions, psig/degrees F	600/650	1,000/650
Condenser, number of passages/flow gallons per minute	single/222,000	single/62,000
Generator, kva/kv	384,000/16,500	96,000/13,800
Exciter, rpm/type	1,800-rpm/motor	1,800/direct connected
Exciter, kw/volts	1,800/500	270/375
Reactor, Steam Plant:		
Design capability (thermal mw)	1,013	240.8
Full load rating (thermal mw)	967	229
Coolant temperature, F, outlet/inlet	675/550	675/600
Coolant flow, pounds per hour	54.5×10^6	21.5×10^6
Number of primary loops	4	3
Fuel (atomic % U-235 in UO_2)	2.5	2.8
Number of fuel elements/control rods	352/61	156/25
Conversion ratio	0.66	0.58
Average burn-up, (thermal mw days/metric ton uranium)	15,000	15,000
Active core size, diameter, height	11′0″×12′0″	7′4″×7′5″
Peak average power ratio, initial/final	2.4/3.5	2.64/3.32
Average specific power, (thermal kw/kilogram)	18.2	16.4
Maximum temperature, F, UO_2 center/cladding surface	3770/850	3980/850
Maximum heat flux Btu/square feet per hour	139,000	150,000
Steam generator and superheaters	U-tube	U-tube
Rating (pounds per hour) at turbine throttle conditions	3.64×10^6	882,000
Reactor building diameter, feet	190	136

* TCDF = tandem-compound double-flow.

Table II. Comparative Power Costs (at 80% Plant Factor)

	Mills/Kwhr		
	305-Emw Coal-Fired	300-Emw OMCR	75-Emw OMCR
Fixed charges	3.30	4.63	8.58
Fuel cost	3.25*	2.45	2.83
Organic makeup		0.63	0.58
Operating and maintenance	0.22	0.41	1.13
Total	6.77	8.12	13.12
Possible future conditions, total	7.24†	7.2‡	12.2‡

* Coal at 35c/million Btu.
† Coal at 40c/million Btu.
‡ Realization of 0.9 mill/kwhr anticipated saving.

1. Reduction in the cost of the polyphenyl material from a present cost of 17 cents per pound appears likely for any substantial volume production. A price of 12 cents per pound for 100 million pounds per year has been estimated by a principal supplier. This would result in a decrease in power cost of 0.19 mill/kwhr. It is also possible that chemical additives will be found which can serve to reduce the rate of radiation decomposition.

2. Fuel preparation and fabrication costs are expected to be reduced by up to 40% by 5 years of development and some experience with plant operation. The resulting savings in power cost would be approximately 0.37 mill/kwhr.

3. It is believed that it will prove practical to utilize APM instead of stainless steel for the fuel element boxes in the core, with a resulting significant improvement in the neutron economy of the reactor. Such a substitution would save about 0.30 mill/kw-hr.

4. It is believed that it will be possible to extend average fuel lifetimes to as much as 20,000 mw-days per ton with a resulting saving of about 0.21 mill/kwhr in the power cost.

5. The reactor and the other plant components are designed with a margin of about 5% over the guaranteed steam flow of the turbine at design temperatures and pressures. If this design margin is realized, the actual full-load capability of the plant would be about 310 emw net under the expected average operating conditions. Such an increase in capacity would reduce the estimated power cost by about 0.16 mill/kwhr.

While the total of these savings is 1.23 mills/kwhr (which would reduce the power cost from 8.12 to 6.89 mills/kwhr), it is not probable that all of such reduction would be fully attained at once or at an early date. However, each of the reductions is considered to have a high degree of probability, and it would seem likely that a reduction in estimated costs approaching 0.90 mill/kwhr is a realistic expectation. The over-all power costs from the 300-emw plant would then be approximately 7.2 mills/kwhr.

Design and construction of a first single-unit 75-emw plant requires an estimated capital investment of $32,150,000, or $427/ekw net. The steady-state fuel-cycle cost for this plant is estimated to be 2.83 mills/kwhr. The addition of fixed, operating, maintenance, and organic make-up costs results in a total power cost of 13.12 mills/kwhr using the same criteria as specified for the larger plant. Since the anticipated cost reductions listed for the 300-emw plant also apply to the smaller plant, it is reasonable to expect that power costs for the 75-emw plant will approach 12.2 mills/kwhr.

COMPARISON WITH 305-EMW COAL-FIRED PLANT

In response to a request from the AEC the Bechtel Corporation compared the economics of a conventional fossil-fueled plant with the large OMCR plant. The study yielded the interesting results summarized in Table II. Power costs of the 75-emw plant are included to indicate the comparative economics of operating a smaller demonstration plant.

A striking observation obtained from comparing the detailed costs of the 300-mw plants is that the equipment costs for the nuclear and conventional plants are identical. The cost for structures is higher for the nuclear plant due almost entirely to the shielding and special building requirements. Indirect and engineering costs and contingencies for the nuclear plant are very much higher than for the conventional plant, reflecting the much less advanced state of the art. However, it is likely that further experience with large nuclear plants will assist in reducing these costs to what might be regarded as a normal level. When this is accomplished, the OMCR capital cost should approach that of a conventional coal-fired plant.

It was noted that the longer construction schedule and start-up time required for the nuclear plant led to significant increases in the cost of interest during construction. Further experience with nuclear plant construction and operation will result in these costs becoming more nearly comparable as well.

The cost of power generation by the conventional plant is very sensitive to the fuel cost on which it is based, since capital charges are relatively low. An increase from 35 cents/million Btu to 40 cents/million Btu, for example, results in an increase of the power generation cost to 7.24 mills/net kwhr. Additionally, an increase in plant factor from the 80% stipulated to 90%, which is typical of base-load operation, is more beneficial to the OMCR than to the conventional plant. When viewed in this light, and when consideration is given to the potential savings which can reasonably be expected in the OMCR fuel and operating cost as outlined above, it is reasonable to expect that, in areas of relatively high and increasing fossil fuel cost, a large OMCR-type unit operating as a base-load plant can become competitive shortly after start-up.

Summary

During the last several years the concept of a reactor plant moderated and cooled by an organic fluid has been established as a feasible approach to nuclear power. Tests at the OMRE and separate experiments have demonstrated that the radiation decomposition rate of the organic represents a nominal additional operating cost item. Measurements of heat transfer rates have shown that the organic fluid can be used under temperature and heat flux conditions of interest over long periods of time. The experimental work and the long-term operation of the OMRE have demonstrated the advantages and disclosed no unexpected difficulties in the potential use of the organic type reactor in central station plants.

Preliminary designs of organic reactor plants have been prepared for sizes ranging from 11.4 emw to 300 emw. A plant designed for 11.4 emw is now under contract for construction in the city of Piqua, Ohio. Other designs using similar operating conditions but for larger power outputs, have been previously prepared. Most recently, an engineering study was conducted by the Bechtel Corporation and Atomics International for the preliminary design of a large organic-cooled and moderated reactor (OMCR) power station. This work was carried out for a 300-emw installation and subsequently modified for a plant of 75-emw output.

The design of the OMCR was based on certain advanced features and operating conditions. Subcooled mucleate boiling was employed as the principal heat transfer mechanism in the core. The fuel elements consisted of assemblies of APM-clad UO_2 rods. The maximum temperature of the fuel rod surface was taken to be 850 F and the design used a relatively high temperature change for the organic passing through the reactor.

Estimates were made of the fuel cycle and capital cost of the 300-emw plant. With escalation and interest during construction, the net cost of power at 80% plant capacity factor was calculated to be 8.12 mills/kwhr. The equivalent figure for a 75-emw plant was 13.12 mills/kwhr. A design was also prepared for the same basic assumptions on a 305-emw coal-fired conventional plant. With fuel costing 35 cents/million Btu, the cost of power for the conventional plant was calculated 6.77 mills/kwhr

Based on the preliminary design of the 300-emw plant, it is apparent that the power cost from an OMCR has not yet reached a value achievable with conventional installations. However, the margin of power cost difference is small and there are a number of factors expected further to reduce the cost of power from the organic-type reactor plant in the future. Based on the present information, it appears that the potential use of this reactor type in central station power plants certainly holds sufficient promise to warrant the construction of prototype units as well as a continuing strong effort directed toward development of the organic reactor technology.

References

1. Organic-Moderated Reactors for Central Station Power, W. E. Parkins, E. F. Weisner. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, 1958 (Jan. 1959 section), pp. 985–93.

2. Summary of Study. "Organic Cooled Power Reactor Study," *Report no. T10-8501*, Bechtel Corporation, San Francisco, Calif., and Atomics International, Canoga Park, Calif., pt. I, Apr. 1959. (available from Government Printing Office, Washington, D. C.)

3. 300 mw Power Plant Conceptual Design. *Ibid.*, pt. 2.

4. Reactor Concept Evaluation. *Ibid.*, pt. 3.

5. 75 mw Power Plant Conceptual Design. *Ibid.*, pt. 4.

6. 300 mw Coal-Fired Power Plant Comparison Study. *Ibid.*, pt. 5.

The Challenge of Universal Mobile Communications

ALAN F. CULBERTSON
MEMBER AIEE

PROCEEDINGS now under way before the Federal Communications Commission are again exploring the possibility that a large block of frequencies might be made available for extension of telephone service to moving vehicles on a large scale. If such an allocation is granted, universal mobile communications will become practically as well as technically feasible for the first time.

Operation of such a system will be a joint venture of some 4,100 operating telephone companies in the United States, of dozens of manufacturers who will supply the system components, and of the various laboratories which contribute long-range system planning for the telephone industry. A peculiarity of this type of service stems from the fact that roughly two-thirds of the territory to be covered by a universal mobile communications system is served by independent telephone companies. This means that a great deal more co-ordination between companies is going to be required than any previous program of similar scope. Thus it is appropriate at this stage to define some of the objectives of a universal mobile communications service and to consider some of the methods by which it can be accomplished.

Experience with the present limited types of 2-way dial mobile telephone service is beginning to uncover some of the more important problem areas which will need to be explored as the service becomes more and more universal in scope. This new approach, along with the "one-way" dial service which has been in operation for a number of years, is providing valuable experience both in the field of control and supervision and also in the field of transmission, where more and more information is being gained about the needs of a commercial grade of telephone service to be provided by radio.

Several important contrasts are immediately apparent between the field of 2-way radio and that of mobile telephone service. The principal difference is the matter of coverage and resulting transmission quality. Telephone subscribers cannot be forced to accept degraded transmission, the use of code words, and special operating procedures. Because of the different approach to coverage required in a universal mobile communications serv-

---

Paper 59-1148, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted June 12, 1959; made available for printing August 25, 1959.

Alan F. Culbertson is with Lenkurt Electric Company, Inc., San Carlos, Calif.

The author wishes to acknowledge the assistance of K. B. Breymayer in working out some of the material in the paper.

ice, it will undoubtedly be found that an altogether new approach in the field of transmission equipment will be required as well.

In order to approach the characteristics of the dial telephone, it will be necessary to find an adequate but inexpensive substitute for the physical wire connection between the conventional telephone and the central office. This physical path is now being made to handle certain essential low-information-content signals both simultaneously and sequentially with speech. These include on-hook and off-hook supervision, dial pulses and ringing current. Performing these same functions via radio is not quite as natural to the medium and considerable research may be required to reduce their cost before they can be applied to the mobile device.

Objectives

Performance Objectives

The objectives to be set for this type of service can be defined very simply; the universal mobile telephone must operate in a manner similar to the wired dial telephone as we know it today, plus any refinements which the dial phone may pick up along the way while the mobile system is under development. These features include the following:

1. Ready availability of dial tone.
2. "Wire-line" transmission standards
   a. Random noise on circuit
   b. Crosstalk
   c. Over-all circuit net loss
   d. Frequency response
   e. In-channel distortion
   f. Echo and singing limitations
3. Switch-hook supervision—on-hook and off-hook
4. Reliable transmission of dial signaling information
5. Positive method of subscriber identification
6. Fully selective signaling
7. Privacy of conversations (so far as party-line type sharing is concerned)
8. Access to direct-distance-dialing network
9. Full-duplex transmission (no push-to-talk button)

Finally, the system must have the ability to connect to the telephone described from any populated area or along any traveled highway.

Design Objectives

A large number of component and subsystem developments are required to meet the performance objectives. All of them must be directed towards a minimum-cost mobile unit. It must also be borne in mind that maintenance of the mobile units will be in the hands of those who now service the automotive broadcast receivers. The amount of extra training and know-how to be absorbed by these people must be kept to a minimum.

A first approximation of the over-all system objectives yields the following as a starting point.

1. Traffic Capacity. A system of 3–5 channels is probably the minimum to start service in any given area, with channel additions based upon an incremental range to be determined by further traffic studies.

2. Coverage Areas will have to be laid out conservatively to achieve the transmission objectives. In a uhf (ultrahigh frequency) system, satellite transmitters as well as receivers may be required for complete coverage. Cross-control of transmitters (similar to the "adjacent-transmitter-lockout" sometimes used in highway mobile service) will be needed to get maximum geographic sharing of the same channel.

3. Transmission Performance to equal "wired" telephone service will have to be along the following lines for at least 50% of the calls for the mobile system:

Frequency response: 300–2,800 cps (cycles per second), 3 db (decibels) points
Noise and crosstalk, combined:
  Toward subscriber: 40 dba (adjusted) maximum, 0 db transmission level
  Toward fixed plant: 35 dba maximum, 0 db transmission level
Subscriber loop loss: 8 db maximum, both directions
Spectrum shift: 20 cps maximum
In-channel distortion: 20% maximum (0 VU talker)

The Spectrum Stretcher

No matter how many channels may ultimately be available in a broadband system, spectrum conservation and principles of good service will demand more efficient channel usage than is possible with one-channel systems. A load-leveling "spectrum stretcher" will have to be used to apply traffic engineering principles to utilization of the radio-derived trunks.

The spectrum stretcher provides access to several different radio channels for each call and marks the specific idle channel in a group which will be used for the next call.

The manner in which the spectrum stretcher can increase the capacity of a group of channels is shown in a typical example; see Fig. 1. If a grade of service equivalent to a 75-subscriber single channel system is assumed, the capacity of 10 channels can be stretched from 750 subscribers in single-channel service to something on the order of 2,100 subscribers by multiple access.

The nominal party-line traffic engineering data from which these figures were originally derived will undoubtedly have to be reviewed in order to develop some ground rules which will apply to the mobile case. It is also certain that our initial objective will be to provide a better grade of service than that afforded by a 75-party line. However, these illustrations do serve to show how basic traffic engineering principles, long applied to the exchange switching problem, can be used to create 28 new channels in the spectrum where only 10 existed before.

Central Office Switching Arrangements

A universal service can be expected to grow until it is a factor to be considered at

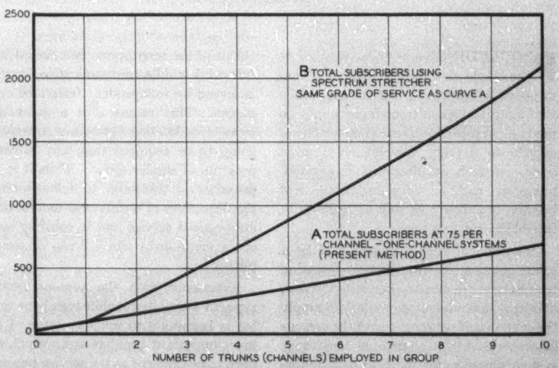


Fig. 1. Increase in the total number of subscribers that can be handled by means of the spectrum stretcher

Vol.
Pt. 1
1959

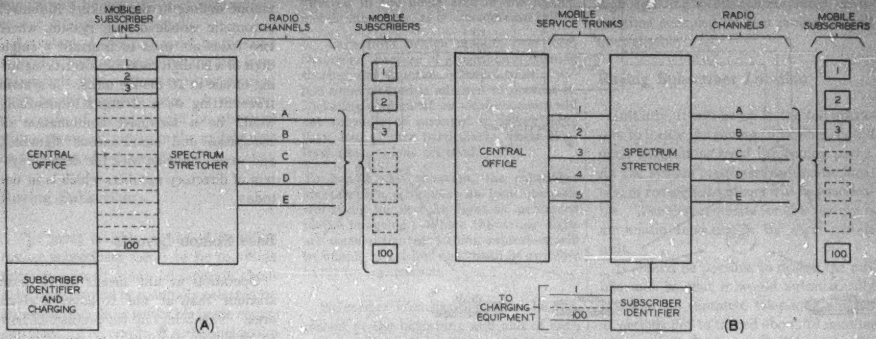

MOBILE SUBSCRIBER LINES
RADIO CHANNELS
MOBILE SUBSCRIBERS
CENTRAL OFFICE
SPECTRUM STRETCHER
SUBSCRIBER IDENTIFIER AND CHARGING
1
2
3
100
A
B
C
D
E
(A)
MOBILE SERVICE TRUNKS
CENTRAL OFFICE
SPECTRUM STRETCHER
1
2
3
4
5
TO CHARGING EQUIPMENT
SUBSCRIBER IDENTIFIER
100
(B)

Fig. 2. Possible switching arrangements for mobile subscribers

A—Line-level arrangement

B—Trunk-level arrangement

every dial-switching exchange in the country. Thus mobile signaling arrangements must be planned to cause the least possible disruption to existing equipment arrangements.

This planning is complicated by the necessity of dealing with panel, crossbar and Strowger step-by-step exchanges as well as with various types of all-relay exchanges. The ultimate system must be able to work with electronic switching as well.

The simplest approach from the standpoint of the central office is the "line-level" arrangement which has been followed in most of the pioneer 2-way dial mobile systems. A block diagram is shown in Fig. 2(A). This equates one mobile station to one land station and assigns a standard directory number to each mobile. Such a system can be adapted to work on practically any type of line circuit available in the central office; 10-party, 4-party, 2-party, and straight lines. Fully-selective signaling the mobile can be achieved to the same extent practiced on the party lines.

Subscriber identification and charging can be accomplished with the line level approach in the same manner as it is for land subscribers. Various arrangements of Strowger Automatic Toll Ticketing (SATT) can be made to work on multi-party lines. Bell System automatic message accounting equipment (AMA) will work directly from one- or 2-party lines.

Widespread use of the spectrum stretcher will influence this picture to some extent. The spectrum stretcher will concentrate a large number of mobile stations into a relatively small number of channels. Thus some equipment savings are possible if no line circuits are provided at the central office. This "trunk-level" approach is illustrated in Fig. 2(B). In this case subscriber identification will have to be accomplished separately by the mobile signaling equipment, possibly as part of the spectrum stretcher.

With line-level termination, selection of the mobile subscriber line will immediately determine the mobile unit to be signaled. With a trunk-level arrangement at least a portion of the mobile subscriber's number will have to be repeated over the mobile service trunk and then translated by the mobile system signaling arrangement. This means more equipment in the mobile signaling portion, but it is common to a large number of mobile units and may have its cost offset by savings in line circuits and in cross connections.

The manner in which a call is placed to a mobile unit with a trunk-level arrangement is shown schematically in Fig. 3. If the mobile unit is associated with the "ROamer 1" central office, then this number might be ROamer 1-2345. Calls for this particular mobile would be directed to the ROamer 1 central office. Within the office, the digits "2345" would be used to select the 2300 trunk group and the last two digits, "45" would be stored.

Upon selection of the marked channel, channel A in the diagram, the digits 45 would be transmitted over it by the mobile signaling scheme. All mobile units in the *2300* group would be monitoring the marked channel, but only one will respond to the digits 45. All other mobiles will scan the remaining channels for the marking indication, and will lock on it to await future calls. The spectrum stretcher can handle only one call at a time for its particular channel group under this scheme, so it is important that the signaling be accomplished in a few milliseconds. It should be fairly easy to operate a high-speed signaling system with the good transmission envisioned for this system.

Mobile Signaling Arrangements

It is evident that a bold new approach to the mobile signaling problem is required to handle an ultimate system with millions

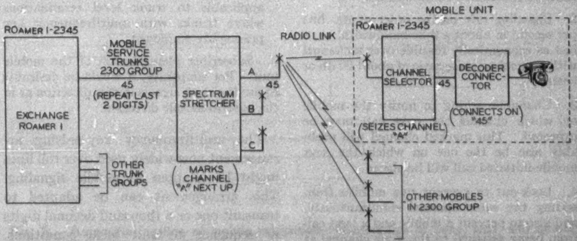

ROAMER 1-2345
EXCHANGE ROAMER 1
MOBILE SERVICE TRUNKS 2300 GROUP
45
(REPEAT LAST 2 DIGITS)
OTHER TRUNK GROUPS
SPECTRUM STRETCHER
A
45
B
C
MARKS CHANNEL "A" NEXT UP
RADIO LINK
MOBILE UNIT
ROAMER 1-2345
CHANNEL SELECTOR
45
DECODER CONNECTOR
(SEIZES CHANNEL "A")
(CONNECTS ON "45")
OTHER MOBILES IN 2300 GROUP

Fig. 3. Illustration of how a land-to-mobile call is placed through the spectrum stretcher

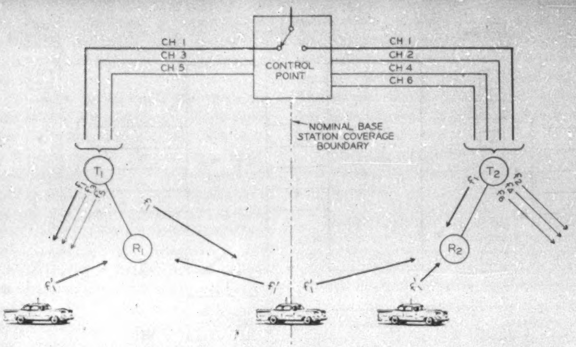


Fig. 4. Possible method of maintaining continuity of conversation as mobile subscriber crosses base station territorial boundary

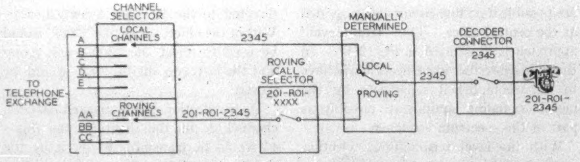


Fig. 5. Possible arrangement of selection and decoding equipment in the mobile unit

of subscribers. Obviously the signaling scheme must be uniform and co-ordinated on a nation-wide basis. To achieve a universal system, the various schemes now using 2-tone sequential, 2-tone simultaneous, 3-tone sequential/simultaneous, etc., will have to be phased out gradually.

With a universal system everyone must be considered a potential roving subscriber and must have a number assignment consistent with this concept.

Some of the attributes of the signaling scheme which will be needed are these:

1. Capacity at the mobile of about 10,000 combinations minimum, expandable on a building block basis to something on the order of 100 million combinations.
2. Signaling rate on the order of 200 bits per second in about a 2-kc bandwidth, or as fast as economically feasible over a channel with a signal-to-noise ratio of about 20 db or better.
3. Channel-marking to notify the mobile on which channel the next call may be expected. This marked channel will probably also be the one on which the next mobile-initiated call will be placed.
4. Lock-out to prevent two mobiles from seizing the same channel simultaneously, and also to prevent a mobile and a land call from being assigned to the same channel at once.
5. A translating scheme which conceivably may take several different forms:
   - a. Choice of the mobile subscriber to be signaled based upon which central office line is being rung. This option is applicable to line level terminations when 1- or 2-party lines are used.
   - b. A choice based upon the combination of the line circuit and the frequency or polarity of the ringing. This is applicable to line-level terminations in offices with multiparty harmonic, decimonic, synchromonic, or polarized fully-selective ringing.
   - c. A decoding of incoming pulses over the mobile service trunk. This option is applicable to trunk-level terminations where trunks with loop dialing or d-c key-pulsing are employed.
   - d. A decoding of incoming tones over the mobile service trunk. This option is applicable to trunk level terminations where trunks with multifrequency key pulsing are employed.
6. Subscriber identification of the mobile unit. For simplicity, it would be desirable to use the same signals in this direction as in the base-to-mobile direction.

The multifrequency key-pulsing arrangement now widely used over toll lines might be adapted to mobile signaling. This arrangement can be adapted to transmit one or a thousand decimal digits in sequence to suit local conditions. Such an arrangement is now being used in mobile dialing in the Lenkurt Mobiltel* automatic mobile dialing system, where two tones are used to transmit a single digit of a 20-digit code for selective signaling of one in 20 mobile units. A system transmitting decimal digits sequentially would be a fortunate combination of sequential and simultaneous signaling, and would coincide with the decimal system of directory numbers which is in use today.

Base Station Layouts

Operation at uhf means more base stations than at vhf to cover a given area. This brings up the problem of how to maintain continuity of conversation to a subscriber who drives across a base station boundary.

One possibility is illustrated in Fig. 4. This arrangement requires that odd-numbered channels normally be marked for use in the territory of base station no. *1*, with even-numbered channels normally used by base station no. *2*.

When a call is initiated while the subscriber is near station no. *1*, he will be placed on an odd-numbered channel; for example channel no. *1*. As he drives within range of receiver *R2*, it will pick up his channel *1* transmission along with *R1*.

After an appropriate time delay, if *R2* is still receiving continuous channel *1* transmission, the land connection can be switched from station *1* to station *2* to continue the conversation. A system such as this will undoubtedly need further refinement to prevent seizure by mobile units from more than one adjacent base station, for example, but it should be workable as a basic principle.

A fundamental problem will be advance determination of base station locations. Expensive field-strength and advance coverage surveys can be justified only in the most heavily populated areas. The equipment used will have to be very thoroughly engineered so that adjustments in coverage area can be made after installation. In a system of this sort the control arrangements which influence seizure of the mobile and marking of the next usable channel will have a decided effect on coverage. Thus adjustments of these functions can possibly have almost as much influence as antenna heights and transmission line losses in a well laid-out system.

It is assumed that all base transmitters will be multiplexed in some form or other. Thus simultaneous multichannel trans-

---

* Copyright applied for.

Vol.
Pt. 1
1959

mission can be accomplished without excessive interchannel crosstalk.

A system of this sort must be planned around a very lower power output from the mobile unit. Thus receiver spacing must be very conservative. A commercial grade of signal should be obtained with a power output from the mobile unit on the order of 1–10 watts.

Roving Subscribers

The most feasible means of handling roving subscribers seems to be to assign them to a small group of separate channels, which would be manually selected by the subscriber as he leaves his home coverage area. The size such a channel group needs to be will have to be determined by actual experience in the field.

Possible arrangement for a roving mobile unit is illustrated in Fig. 5. In this arrangement a mobile unit with the full directory number of 201-ROamer 1-2345 is assumed.

Local channels to which this mobile is assigned all terminate in the ROamer 1 central office in the 201 numbering plan area. Thus when the mobile is operating on local channels, the local call selector can be permitted to respond to any call which transmits the digits 2345.

When the subscriber switches manually to the roving channel group, he increases the discrimination of his selector by cutting in ahead of it a device which will only connect through calls in which the first digits are 201-RO1.

In case there are subscribers who have no use for the roving feature, the roving call selector should be designed so that it can be added to the mobile unit on a building block basis. In large metropolitan areas a selector may be required which will respond to 7 digits even though the subscriber does not roam far enough to require the total of 10 digits. For this application, the *RO1* section should be made available separately from the *201* section.

The mobile unit will have to be arranged to transmit its complete 10-digit identification for all calls it initiates, both local and roving. The following conditions are possible:

1. Local subscribers' calls placed over local channels: mobile identification equipment will verify that the first 6 digits correspond to the local area, and then absorb them. AMA or SATT equipment records only the last four digits, adequate for billing purposes to local mobile subscribers.

2. Calls from foreign mobiles over roving channels: roving channels come into special ticketing equipment which must be arranged to record all 10 digits. Bills are directed to the home area of the foreign mobile subscriber, as in a credit-card call.

3. Calls from foreign mobiles over local channels: because of geographical channel sharing, this situation is bound to exist unless some method is adopted to prevent it. Ticketing equipment on local channels will not normally be arranged to record more than four digits, particularly where line-level terminations are used.

To handle this situation, the subscriber identification equipment on local channels will check the first six digits as mentioned in the foregoing. Where the wrong digits are transmitted, the calling subscriber will be unable to get dial tone until he switches to his roving channels.

Subscriber identification must be furnished at the beginning and end of each transmission. This will provide additional verification in case of fades or other interruptions to transmission. With the high signaling speeds contemplated, channel time occupied by this procedure will be negligible.

Roving Subscriber Location

Initially, it will be up to the toll operators to locate the roving subscriber, based upon suggestions from the calling party. As the system progresses automatic location of roving subscribers will be practicable. The first requisite for this feature is automatic reporting-in by each mobile unit.

It should be possible to design the mobile unit so that it would automatically transmit its complete telephone number at periods not to exceed about 10 minutes whenever the selector switch is in the "Roving" position.

The base receiver picking up such trans-

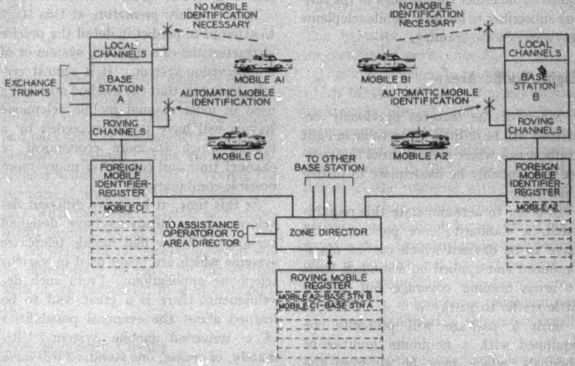

NO MOBILE IDENTIFICATION NECESSARY
LOCAL CHANNELS
EXCHANGE TRUNKS
BASE STATION A
MOBILE A1
AUTOMATIC MOBILE IDENTIFICATION
ROVING CHANNELS
MOBILE C1
FOREIGN MOBILE IDENTIFIER-REGISTER
MOBILE C1
NO MOBILE IDENTIFICATION NECESSARY
LOCAL CHANNELS
MOBILE B1
BASE STATION B
AUTOMATIC MOBILE IDENTIFICATION
ROVING CHANNELS
MOBILE A2
TO OTHER BASE STATION
FOREIGN MOBILE IDENTIFIER-REGISTER
MOBILE A2
TO ASSISTANCE OPERATOR OR TO AREA DIRECTOR
ZONE DIRECTOR
ROVING MOBILE REGISTER
MOBILE A2-BASE STN B
MOBILE C1-BASE STN A

Fig. 6. A method for automatic locating of mobile subscribers

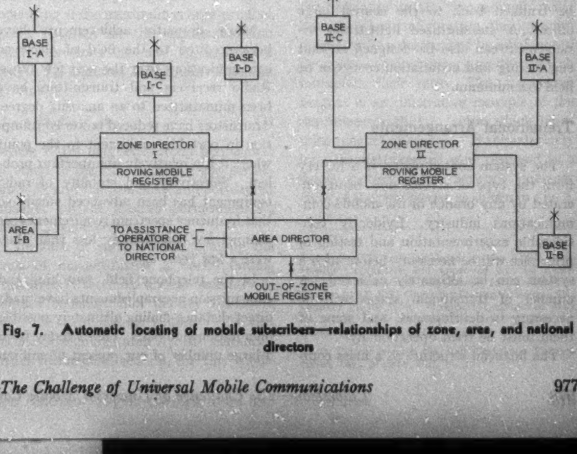

BASE I-A
BASE I-C
BASE I-D
BASE II-C
BASE II-A
ZONE DIRECTOR I
ROVING MOBILE REGISTER
ZONE DIRECTOR II
ROVING MOBILE REGISTER
AREA I-B
TO ASSISTANCE OPERATOR OR TO NATIONAL DIRECTOR
AREA DIRECTOR
BASE II-B
OUT-OF-ZONE MOBILE REGISTER

Fig. 7. Automatic locating of mobile subscribers—relationships of zone, area, and national directors

missions will record them in its identifier register and will feed it to a "roving mobile register" a high-speed memory device, as shown in Fig. 6.

Calls for roving mobiles are then dialed into a "Zone Director" which is associated with the roving mobile register. The zone director would then complete the call through the base station from which the mobile unit was last reported. The roving mobile register will erase all entries stored in it after 10 minutes have elapsed. If the mobile unit is not heard from again, it will be assumed to have gone back to local channel operation or to have left the zone.

Area registers, as shown in Fig. 7, will eventually have to be added to correlate information on out-of-zone mobiles which will be relayed by the zone directors. Ultimately, a national director can be added and no assistance should be required from either the operator or the calling subscriber to locate a mobile telephone anywhere in the country.

Light Traffic Areas

Many of the features previously described will be of little or no value in light traffic areas, where revenue from this service will probably be inadequate to justify their cost.

In order to accommodate this requirement, a standard "core package" will have to be devised which can be mass-produced and applied on a random basis to areas needing coverage but having little traffic to justify it.

Such a package will probably be equipped with a minimum number of channels, which may be all from the roving group. Because of the light traffic density, complex cross-control features can be eliminated. Subscriber identification information can possibly be trunked back to the nearest large office. A standardized light-traffic arrangement can also be designed so that engineering and installation costs can be held to a minimum.

Transitional Arrangements

The system contemplated is a far cry from the type of service now being operated by any branch of the mobile communications industry. Evidently considerable experimentation and testing of principles will be necessary before such a system can be ultimately achieved. A number of transitional stages will be necessary in development, and some of them must be tried operationally.

The financial structure of a mass communications industry dictates that new services integrate as smoothly as possible on a planned evolutionary basis. Thus existing plant can be amortized in an orderly fashion without undue penalty to the new service.

Unfortunately, the nature of frequency allocations is such that if a universal service is to be provided, it is almost certain to be in a band which will require a five-to-one jump in frequency for most of the sets now in use. There can be no gradual transition for radio equipment under these conditions, and entire systems will have to be cut over abruptly when the new equipment goes in service.

There is much about the control equipment which can be reused from one system to another. Our objectives at this stage should be to think of control equipment now under design as being adaptable as the first stage in the trend towards a universal mobile communications service.

It is certainly premature at this stage to attempt to predict in detail the precise characteristics of a signaling system or of a supervision system. It is almost certain, however, that the method now being used in most manual mobile telephone service will have to be succeeded by a system which is more economical of channel time and capable of many more possible combinations.

At this time, at least five different development organizations have designed versions of 2-way dial mobile telephone systems which are being used in various telephone applications. From such developments there is a great deal to be learned about the eventual possibilities of a universal mobile system. Ultimately, of course, one standard will have to be agreed upon and the universal system will have to be organized.

Conclusion

Many dramatic achievements have been credited to the field of electrical communications over the past few years. Radio receivers and transmitters have been miniaturized to an amazing degree. Transistors have reduced power consumption in portable equipment to the point where it is a relatively unimportant problem. Selectivity and stability of radio equipment has been advanced steady so that frequency spectrum requirements per channel are considerably less than they were a few years ago.

In the telephone field, switching and transmission accomplishments have made direct-distance-dialing ultimately possible on a nationwide basis. Even today from a large number of our present 65 million telephones, many other telephones in the country can be reached in a matter of seconds with just a few pulls on the dial.

Table I. Total Number of Subscribers Served per Trunk Group

No. Trunks Accessible to Each Vehicle	Number of Subscribers Served		
	Holding Time = 100 Sec. Delay = 25 Sec.	Holding Time = 80 Sec. Delay = 5 Sec.	Comparable to Current One-Channel Service
1	12	5	75
2	43	15	291
3	75	59	469
4	110	93	688
5	144	133	900
6	182	170	1,137
7	210	210	1,350
8	245	252	1,574
9	280	292	1,799
10	310	333	1,992

The high standards of performance required from both people and equipment to permit this accomplishment are staggering to contemplate. Telephone service of this type demands a degree of reliability unheard of in other applications of electronics. The problem of transmission between any random pair of telephones is much more challenging than any specific point-to-point communications problem.

Growth of the wired telephone system to this point has been dependent upon obtaining a right-of-way for the poles, wires and cables which comprise the fixed transmission paths between subscribers. Growth of the unwired telephone service can begin as soon as an adequate right-of-way is obtained through the frequency spectrum to reach the subscribers whose telephones can not be anchored in place.

Appendix I. Traffic Engineering Considerations

By applying standard formulas used in the planning of dial central offices,[1] one may derive various values for the number of trunks needed to handle a given number of subscribers. If traffic in a certain office is known to have an average holding time of 100 seconds, and it is engineered to an average delay of 25 seconds, the figures in first column of Table I indicate the total number of subscribers that can be served by using from one to ten trunks.

The figures in the second column give the same information if the holding time is assumed to be 80 seconds and it is engineered to an average delay of 5 seconds.

It is known that mobile service today is based upon common sharing of a channel by as many as 75 subscribers. In private systems the figure will run much higher than 75. Therefore as a starting point it is appropriate to speculate about the service

Vol.
Pt. 1
1959

capabilities based upon performance no worse than that obtained today with single-channel systems. Fitting a curve to the figures of the first and second columns yields the figures in the third column. These figures are plotted in Fig. 1 of the main body of this paper.

It is quite likely that these heavy subscriber densities will upset the previous rules. There is certain to be a "mobile traffic factor" of some kind which reflects the smaller percentage of time that the subscriber spends in his vehicle, the possibility that he has driven out of the area, and the possibility that the telephone has been turned off. Only time and experience will determine the proper criteria to follow for this type of service.

References

1. Delay Probability Charts for Telephone Traffic Where the Holding Times are Constant, Imre Molnar. *Engineering Notes 2032*, Automatic Electric Company, Chicago, Ill., May 1, 1952.

2. Application of the Theory of Probability to Telephone Trunking Problems, E. C. Molina. *Bell System Technical Journal*, New York, N. Y., vol. 6, July 1927, pp. 461–94.

3. Probability in Engineering, E. C. Molina. *Electrical Engineering*, vol. 54, Apr. 1935, pp. 423–27.

4. Application of Poisson's Probability Summation, Frances Thorndike. *Bell System Technical Journal*, New York, N. Y., vol. 5, Oct. 1926, pp. 604–24.

5. Telephone Traffic—Terminology and Calculations. "Telephone Engineering and Construction Manual," Rural Electrification Administration, Washington, D. C., section 505, Issue no. 3, Apr. 1956.

6. Telephone Traffic—Dial Central Office Equipment Switch Quantities. "Telephone Engineering and Construction Manual," Rural Electrification Administration, section 510, Issue no. 4, Apr. 1956.

Relation Between Permanent Magnet Configuration and Performance

P. P. CIOFFI
MEMBER AIEE

THE NECESSITY for minimizing magnet weight has raised questions regarding the most efficient permanent-magnet shape for the magnetic circuit of air-borne magnetrons. The bowl-shaped permanent magnet was introduced, presumably because of its greater efficiency, to replace the conventional U-type permanent magnet.[1] As far as is known, however, no direct evidence has been presented to support this view, and the purpose of this investigation was to determine by magnetic-circuit analysis the relative merits of the two permanent-magnet configurations.

The permanent-magnet structure in question consists of two bowl-shaped Alnico V magnets with their brims in contact and with axial cylindrical pole faces in contact with the magnetron cylindrical pole pieces; see Fig. 1(A). In this manner the volume of the permanent magnet is circularly distributed except for an open wedge for bringing out the output waveguide. The magnet path length is the distance measured along the mean flux path from pole to pole and is determined by the mmf (magnetomotive force) necessary to develop the required air-gap field.

The more common permanent-magnet structure consists of two U shapes about oppositely disposed faces of the magnetron pole pieces; see Fig. 1(B).

The volume of permanent-magnet material and its performance in the magnetic circuit are compared for the two configurations. However, an appreciable part of the total weight of the magnetic-circuit assembly, which must be kept as small as possible for air-borne applications, is in the provision for the cooling and packaging of the magnetron with the permanent magnet. This must also include a nonmagnetic cover to protect the permanent magnet against possible contact with magnetic material and consequent degradation of air-gap field. The excess weight of the magnetic circuit will be determined, in part, by the surface area of the permanent-magnet configuration.

The Magnetic Circuit

The permanent magnet must be proportioned to develop an mmf across its pole faces equivalent to the drop in mmf in the air-gap and pole pieces. A magnet operating at the maximum energy-product point $(\mathbf{B}_m,\mathbf{H}_m)$ on the demagnetization curve, develops the mmf $\mathbf{H}_m,L_m$ with a length L_m and area at the neutral section given by

$$A_m=(\mathbf{H}_mL_m)P/\mathbf{B}_m \qquad (1)$$

where P, the magnetic-circuit permeance, is equal to the sum of the air-gap permeance P_g and the pole piece and permanent magnet surface leakage permeances P_p and P_m respectively.

Because of surface leakage along the axis of magnetization, a corresponding variation of sectional area is necessary to obtain constant flux density operation over the entire magnet volume. The variation of sectional area A, with distance l from the neutral section, for this condition of operation, is illustrated in Fig. 2 for both the U- and bowl magnets which were examined in this study. If f is the average height under the curves of Fig. 2, the average sectional area is fA_m and the magnet volume is

$$V_m=(\mathbf{H}_mL_m)^2fP/\mathbf{B}_m\mathbf{H}_m \qquad (2)$$

The quantity (fP) represents average magnet circuit permeance P_a.

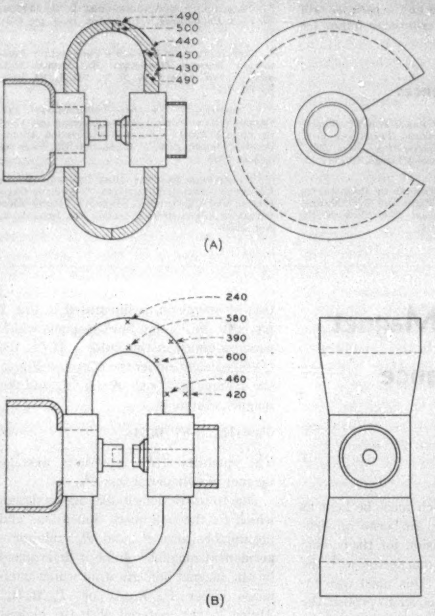
Due to restrictions in magnetron design which fix the pole piece dimensions and presumably also P_g and P_p, minimum permanent-magnet volume is determined by the magnet configuration which minimizes either $P_m/\mathbf{B}_m\mathbf{H}_m$ or $P_a/\mathbf{B}_m\mathbf{H}_m$. Alnico V, the material with the highest energy product commercially available, satisfies one condition for minimum magnet volume. Since surface leakage permeance is proportional to surface area, a magnet configuration with circular cross sections, Fig. 3, and hence minimum surface area, would satisfy the second condition.

The concept presented here, however, is that permeance and energy product are mutually dependent and that least permeance and maximum energy product are incompatible except in ellipsoids of revolution of appropriate dimensional ratio. For any other configuration, the optimum magnet will be determined by a balance of permeance against energy product, which minimizes their ratio. The magnetron magnet is an illustrative example of the dependence of the two factors which determine magnet efficiency. If the relatively long magnet required were disposed as a single curved shape between coplanar faces of the pole pieces, as shown in Fig. 3,

---

Paper 59-1120, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted May 14, 1959; made available for printing July 17, 1959.

P. P. Cioffi is with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.



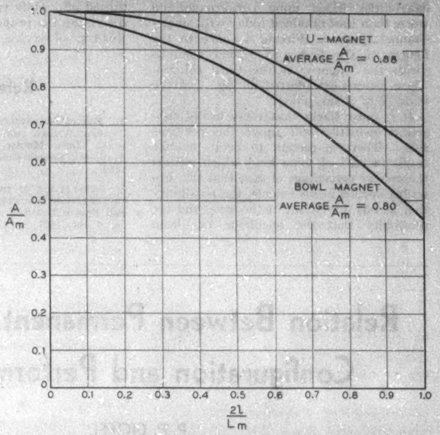


Fig. 1 (left). Magnetic circuit with (A) bowl magnet and (B) U-magnet configuration. Operating field in the permanent-magnet material as indicated

Fig. 2 (above). Variation of magnet sectional area with distance from neutral section under constant flux density operation

Vol.
Pt. 1
1959

due to unequal flux paths measured from the pole piece equipotential surfaces, the permanent-magnet material would be nonuniformly magnetized and the greater the magnet thickness the greater would be the degree of nonuniformity of magnetization. Assuming operation of the axial path at the optimum point on the demagnetization curve, material on either side would operate above or below this point, i.e., at lower energy product, and average operation below maximum would depend upon the degree of nonuniformity of magnetization.

With a single magnet of circular sectional areas, the condition for minimum surface permeance, the greatest degree of nonuniform magnetization would prevail and operation would be substantially below optimum. By distributing the permanent-magnet material between two magnets, energy product operation would improve as a result of the decrease in radial thickness and consequent increase in uniformity of magnetization. The maximum degree of uniformity of magnetization would be obtained by distributing as much as possible of the permanent-magnet material about the pole piece axis, as in the bowl magnet configuration. The expediency of redistributing magnet volume to improve energy product operation is accompanied, however, by an increase in surface area and leakage permeance.

Magnetic Measurements

To arrive at some assessment of the relative performances of the two types of magnetic circuits for the same practical application, detailed measurements were made on a sample of each type; these data are given in Table I. In this comparison, the magnet volumes differ by only 1.4% and the pole pieces are dimensionally alike except that the contact surfaces with the magnet pole faces are either cylindrical, Fig. 1(A), or rectangular, Fig. 1(B). With comparable permanent-magnet material volumes and pole piece dimensions, differences in magnetic-circuit behavior can be ascribed mainly to differences in configuration of permanent-magnet material.

Magnetic measurements were made with the recording fluxmeter with an estimated error of less than 0.5%.[2] The flux through various sections of the magnet was measured with a single-turn search coil. Measurements on the U-magnet configuration were made by slipping the coil from a given section of one magnet to the corresponding section on the second magnet. Measurements on the bowl magnet were made by removing the coil from the magnet after disassembly of the circuit. No error is introduced by the relatively long interval required for measurements because of the very large time constant of the fluxmeter. The curves of Fig. 2 were derived from these results.

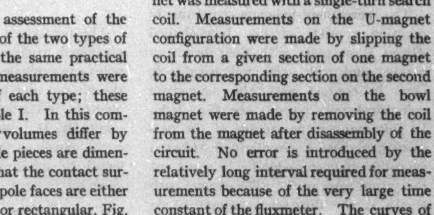


Fig. 3. Magnetic circuit with single magnet of circular cross sections and minimum surface

Table I. Comparison of Magnet Configurations

Magnet Configuration	Weight, Pounds	Volume, Cm³	Air-Gap Field, Oe	Mmf, Oe-Cm	Air-Gap, Cm		Average B_mH_m, Gauss-Oe	Permeances, Cm					P_a
					Nominal	Apparent		P	P_g	P_p	P_m	P_a	B_mH_m
U	4.50	278	7,400	5,650	0.636	0.765	4.49×10^6	44.6	1.24	26.8	16.6	39.1	$8.7\times10^{-}$
Bowl	4.56	282	7,425	5.575		0.750	4.15	47.0		20.1	25.7	37.6	9.1

Cm = centimeters.

Measurements of magnetic-circuit mmf and field intensity in the permanent-magnet material were made with Chattock potentiometer coils.[3,4] Permeances were determined from mmf and flux measurements and average energy product operation of the permanent magnet was determined from equation 2.

The air-gap field $\mathbf{H}_g$ was measured with a small search coil midway between the inner and outer radii of the pole pieces at the center of the air gap. The inconsistency between air-gap field and mmf in these measurements is due to a small discrepancy in air-gap length between the two magnetic circuits. This is evident as an inequality in apparent air gap, given by $\mathbf{H}_mL_m/\mathbf{H}_g$. The apparent air gap is appreciably larger than the nominal dimension because of magnetic saturation in regions of the pole pieces.

Discussion of Results

The striking observation in the results presented in Table I is that the total magnetic-circuit permeance P is only 5% greater for the bowl magnet configuration, and the significant difference between the two magnetic circuits is in the division of permeance between the magnet and pole piece surfaces. The pole piece surface leakage permeance P_p is 33% higher for the U-magnet configuration because the relatively smaller contact surface with the magnet pole faces leaves large exposed areas at full magnetic potential difference. On the other hand, the bowl magnet configuration has 35% greater magnet surface leakage permeance P_m because of its greater surface area, especially in the regions of maximum potential difference. These results indicate that total magnetic-circuit permeance is, for all practical purposes, independent of permanent-magnet configuration and only its division between magnet and pole pieces is affected by configuration. This is to be expected since the major contribution to permeance derives from the polar regions, and the magnet configuration determines whether the predominance of this surface resides in the magnet or in the pole pieces.

The second point of comparison is energy product operation of the permanent-magnet material. The U-magnet configuration operates with approximately 8% higher energy product but, since average permeance is also slightly greater, the ratio $P_a/B_m\mathbf{H}_m$ is not very dissimilar for the two configurations. Average energy product operation, however, is substantially below the characteristic value of 5×10^6 gauss-oe (oersteds) for Alnico V in both configurations. This can be accounted for in the bowl shape by incomplete development of anisotropic characteristics in the direction of final magnetization because of inadequate processing. The behavior of the U-magnet, on the other hand, can be accounted for by nonuniform magnetization and consequent operation at a level of energy product below the intrinsic value. An indication of the relative degree of uniformity of magnetization in the two magnet configurations is illustrated by the difference in field intensity between the outer and inner magnet edges, ranging from approximately 2% to 12% in the bowl magnet, Fig. 1(A), and 8% to 60% in the U-magnet, Fig. 1(B).

The anisotropic characteristics of Alnico V are developed by cooling, after heat treatment, at a controlled rate through the Curie temperature in a magnetic field of appropriate intensity applied in the direction of actual magnetization of the magnet in its magnetic circuit. A misalignment of the processing field will prevent the development of maximum permanent-magnet capability in the direction of magnetization. The processing field for bowl shapes is developed with pancake coils, and inaccurate field shaping by this method may account for the characteristics observed. The U-magnet configuration, however, is better adapted for processing because the magnetic field can be accurately directed with a solenoid of appropriate shape around the magnet. The relatively high average energy product actually obtained, compared with the maximum possible, is indicative of the effectiveness of processing in this shape.

Since magnetic-circuit permeance is practically independent of magnet configuration, energy product operation is the predominant factor in magnetic-circuit efficiency. The U-magnet configuration is potentially capable of operating with the optimum characteristics of the permanent-magnet material but this capability is not realized when the radial thickness is large because of nonuniform magnetization and consequent average operation substantially below optimum energy product. However, since total magnetic-circuit permeance is independent of magnet surface area, the magnet thickness can be made as small as necessary to take advantage of uniform magnetization and maximum utilization of the permanent-magnet characteristics.

References

1. ELECTRON DISCHARGE DEVICE WITH TOROIDAL PERMANENT MAGNET, Paul W. Crapuchettes. *U. S. Patent no. 2,787,728*, Apr. 2, 1957.

2. A RECORDING FLUXMETER OF HIGH ACCURACY AND SENSITIVITY, P. P. Cioffi. *Review of Scientific Instruments*, New York, N. Y., vol. 21, 1950, pp. 624–28.

3. ON A MAGNETIC POTENTIOMETER, A. P. Chattock. *Philosophical Magazine and Journal of Science*, London, England, vol. 24, 1887, pp. 94–96.

4. THE MAGNETIC POTENTIOMETER STUDY OF PERMANENT MAGNETS, L. F. Bates. *Philosophical Magazine and Journal of Science*, vol. 36, 1945, pp. 297–318.

◆

Discussion

M. J. Schindler (Radio Corporation of America, Harrison, N. J.): When Faraday introduced the concept of closed lines of induction, he predetermined the evolution of magnetic engineering up to the present. In spite of the fact that the induction **B** has no physical meaning, while the magnetization (or intrinsic induction) **I** does, we still insist on using the former and not the latter. The field in a magnetic object is given by the sum of the external field and the demagnetizing field, which is a function of the magnetization **I** and not of the induction **B**. Therefore, the use of a **B-H** plane in the presence of external fields will even yield incorrect results.

In the case discussed by Mr. Cioffi, use of the **I-H** concept would simplify the issue inasmuch as the distinction between pole piece and magnet stray flux disappears. A given configuration with a given distribution of surface potential will create a certain field configuration and the local magnitude of this field will determine the local operating point of the magnet. Such a static concept of permanent-magnet devices will not, of course, permit thinking in terms of an electrodynamic analogy and therefore it does away with load lines, lumped permeances, magnetic "circuits" and, generally, "Ohm's law of magnetism." On the other hand, it indicates the usability of electrolytic tanks, resistance paper and similar procedures of field analysis, and should provide the design engineer not only with a better understanding of the problem, but also with a more accurate solution.

Recently, I conducted some experiments related to the design of periodic magnetic focusing stacks for traveling-wave tubes. They clearly indicated the superiority of the static concept over the generally used "electrodynamic" design approach, yielding very accurate results.

P. P. Cioffi: Although Dr. Schindler's remarks are thought-provoking, it should be pointed out that no ambiguity will arise in the solution of magnetic-circuit problems if the distinction and relationship between **I** and **B** are kept clearly in mind. An important aspect of problems of this kind, for instance, is the relationship between energy stored in the air gap and the energy product developed by the permanent-magnet material. This is most conveniently expressed in terms of the product of the **B** and **H** co-ordinates of the operating point of the material on the demagnetization curve. While the static concept and the use of the electrolytic tank and similar design aids are useful in magnetic-circuit problem investigations, such methods are of dubious value in the analysis of actual magnetic-circuit configurations compared with the direct method presented in this paper.

Use of High-Speed Digital Computers to Study Performance of Complex Switching Networks Incorporating Time Delays

Y. N. CHANG
ASSOCIATE MEMBER AIEE

O. M. GEORGE
NONMEMBER AIEE

THE two-valued characteristic of a switching network (i.e., the "make" or "break" of a contact, the presence or absence of a voltage at a node, or the activation or deactivation of an electric or mechanical control element) makes it possible to apply the principles of Boolean algebra, an algebra of logic, to the design and analysis of the network. The first important contribution to switching algebra was made by Shannon in 1938.[1] The fundamental concepts, postulates, and theorems of switching algebra are well treated in many books; consequently they will not be presented in this report. Most of the terminology used in this presentation can be found in reference 2. The concept of transmission was chosen for this report on the application of a high-speed digital computer to the analysis of switching networks, although the concept of hindrance could be used equally well.

Switching algebra has been applied to the design and analysis of static switching networks, but sequential switching networks incorporate make and break time delays that introduce hazards if the sequence of events is not properly controlled. The bookkeeping involved in maintaining a time history of each logical element makes calculation very laborious, time consuming, and subject to error. The modern electronic digital computer has three advantages over the engineer in this phase of the problem: (1) it has an accurate memory, (2) it does arithmetic at a fantastic rate, and (3) it has no aversion to tedious and repetitious computations. An extension of the switching algebra notation has been developed by the authors to facilitate the analysis of sequential switching networks on a digital computer. The computer program described in this paper enables the engineer to study the dynamic performance of a complex switching network by including the actual time delays of the logical elements. Therefore, he can check performance, improve the design, and discover sneak circuits caused by improper timing, relay racing, and incompatible conditions, without having to construct and activate the system.

The high-speed digital computer will tabulate a complete time history of a switching network comprising hundreds of logical elements in a minute or less. Thus the engineer can explore the switching network efficiently for input combinations which may not have been considered in the original design. A complex switching network usually consists of several seemingly independent parts. With the aid of a computer, a combined system performance study becomes feasible and economical.

Mechanization of Switching Network for Digital Computer Programing

ELEMENTS, VARIABLES, AND STATES

Digit *1* is used to represent the "1" state (on, closed, energized, activated, etc.), and digit *0* represents the opposite "0" state. In a network, any physical element which has two states is defined as a "logical element." An on-off switch, a relay, the presence or absence of voltage on a control bus or node, an open-and-shut valve, or any other 2-state device is a logical element. A device which has more than two states must be resolved into *n* logical elements. For example, if the device is an *n*-position switch, each of the *n* logical elements is in state *1* when the switch makes contact at its own position, and state *0* when the switch is not making contact at this position. Additional examples are given in Appendix I. A system comprising logical elements is a logical system (reference 2, page 7). To permit the use of a digital computer to study a complex system, all the logical elements in the system are identified by numbers.

Input elements are those logical elements whose states are independent of other elements. Their states are assigned by the investigator from time to time as conditions of system operation.

Dependent elements are those logical elements whose states are to be determined from a Boolean function of other elements. These other elements are called variables. A variable is given the same identification number as its associated

Paper 59-1125, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11-16, 1959. Manuscript submitted June 15, 1959; made available for printing July 22, 1959.

Y. N. CHANG and O. M. GEORGE are with North American Aviation, Inc., Downey, Calif.

This paper describes a digital computer program developed under U.S. Air Force contract no. *AF33(600)-28469*.

Vol. 7
Pt. 1
1959

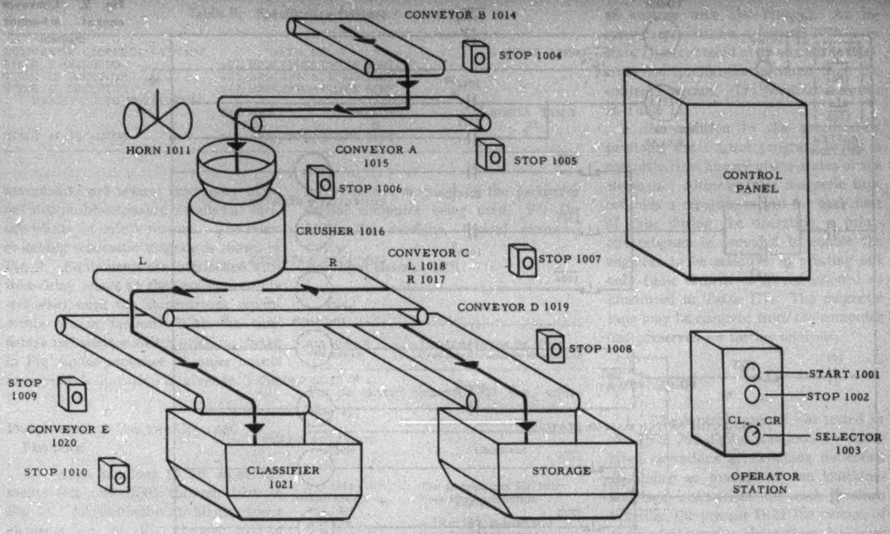


Fig. 1. Conveyor system diagram

logical element. For instance, a transmission function from control voltage to ground for a relay coil may be used to determine the state of this element. When this function becomes *1*, the element is energized; when this function becomes 0, the element is de-energized. However, the contacts of this relay are variables, because they appear in the transmission functions of other dependent elements.

A complemented variable assumes the opposite state of its associated element. A good example of this is a normally closed contact of a relay. When the relay is energized, state *1*, the contact is open, state *0*, and conversely. The complementation is represented by a bar over the number designation of the element. A bar is also used over an expression to represent the complementation of that expression as in conventional Boolean algebra.

Time-Delay Elements

In addition to the sequential nature of the input conditions of a logical system, the elements themselves are frequently the cause of sequential events. Two notations are introduced here to represent time delays of an element: the α-delay is written as a numerical superscript and the β-delay as a numerical subscript on the right-hand side of the number designation of the variable. Each variable may have its own set of α and β values, although the variables are associated with a common element (e.g., a relay may have several different time-delay contacts).

The α-delay is defined as the integer units of time delay between a change of state from *0* to *1* of the element and the change from *0* to *1* of the variable, or *1* to *0* for a complemented variable. The β-delay is similarly defined as the time between a change of state from *1* to *0* of the element and the change of state of the variable. The omission of α and β implies that the change of state of the variable occurs instantaneously with that of the element (i.e., no time delay). The unit of time may be minutes, seconds, milliseconds or any other convenient unit as long as α and β are integers.

An example of the mechanization of a system is given following the description of the digital computer control program.

Functional Description of Digital Computer Program

Control Program

A basic control program was written for the International Business Machines Corporation (IBM) *704* digital computer to check the compatibility of a switching network and the sequence of events directed by (1) input conditions entered at intervals during the simulated operation, and (2) the time delays of the elements.

The program essentially solves a set of simultaneous Boolean equations which describe the logical system by expressing each dependent element as a function of other variables so that the state of each logical element is determined at any time. At the option of the engineer, the state is compared with a specification (the expected state at that time entered as data) and the results of the comparison are printed. If the specification is met, the control program advances a time counter one unit and continues the simulated operation. The engineer may direct the program to continue if the specification is not met; otherwise it stops. At the termination time, entered as data by the engineer, the history of the states of the elements has been written on a magnetic tape, one record for each unit of time. Provision is made for entering a brief description of the conditions of the operation. This description is written at the beginning of the magnetic tape to identify the performance tape when it is removed from the computer and stored for subsequent study.

The simulated operation of a system starts at *0* time. The control program determines the equilibrium state of each logical element by an iterative solution of

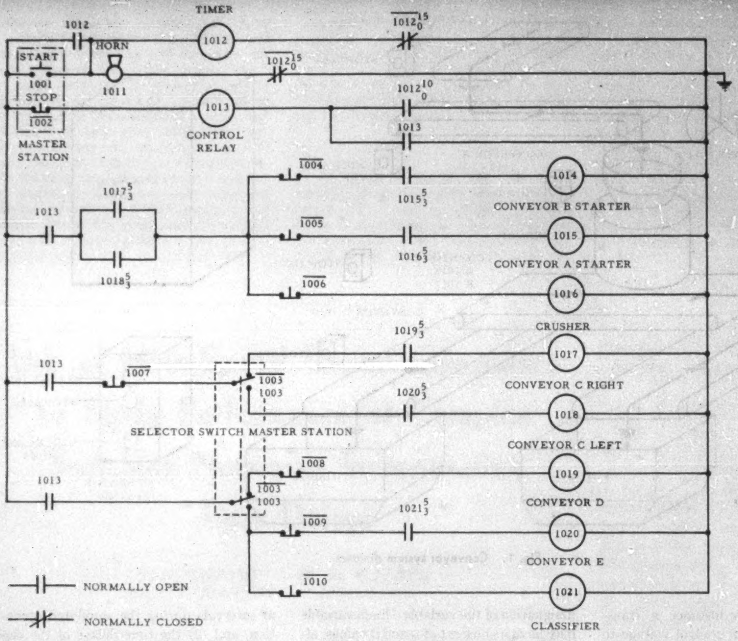

TIMER
1012
HORN
START
1001
STOP
1002
MASTER STATION
1011
1012
1013
CONTROL RELAY
$\overline{1012}{}^{15}_{0}$
$\overline{1012}{}^{15}_{0}$
1012^{10}_{0}
1013
$\overline{1004}$
1015^{5}_{3}
1014
CONVEYOR B STARTER
1013
1017^{5}_{3}
1018^{5}_{3}
$\overline{1005}$
1016^{5}_{3}
1015
CONVEYOR A STARTER
$\overline{1006}$
1016
CRUSHER
1019^{5}_{3}
1017
CONVEYOR C RIGHT
1013
$\overline{1007}$
$\overline{1003}$
1003
1020^{5}_{3}
1018
CONVEYOR C LEFT
SELECTOR SWITCH MASTER STATION
$\overline{1008}$
1019
CONVEYOR D
1013
$\overline{1003}$
1003
1021^{5}_{3}
$\overline{1009}$
1020
CONVEYOR E
$\overline{1010}$
1021
CLASSIFIER
NORMALLY OPEN
NORMALLY CLOSED

Fig. 2. Conveyor control schematic diagram

the set of Boolean equations, and records the time and state. A time counter location is increased by one unit and the equilibrium states are redetermined. At each time, the program checks for new input conditions; thus, the states of input elements remain constant until the engineer enters new data. A time-delay subroutine applies the α- and β-delays of a variable to the recorded time of change of state of the associated element to determine the effective state of the variable at each unit of time. Debugging routines print diagnostic comments for indeterminate and incompatible states, or the improper entry of input states and specifications, and stop the simulated operation.

Boolean Equations Program

The set of Boolean equations for the dependent elements of the network are coded for the digital computer and then assembled in machine language to form a subprogram compatible with the basic control program. The control program is permanent; only the Boolean program need be reassembled for a new network. Since the Boolean program is assembled in relocatable form, the equations may be changed directly in machine language without reassembling. If extensive additions are made to the network, it is preferable to reassemble the Boolean program.

Performance Print Program

At the conclusion of the simulated operation, the print deck may be loaded to print selected records from the performance tape, or the tape may be removed and stored for subsequent printing. The print program prints the system description stored on the tape. Then it searches the tape and prints the states of all the input and dependent elements at each time selected by the engineer and entered as data.

Application of Program to a Simplified System

The following example was chosen to illustrate the procedures for using the control program and to show the input data and the printed results of the analysis of a switching network.

Description of the System

The physical arrangement of the system is shown in Fig. 1. Conveyor *B* transports ore from a loading area to the plant site. Conveyor *A* transfers rough ore from conveyor *B* to the crusher. After being crushed, the fine ore can either be directed through conveyors *C* and *D* to a storage bin or through *C* and *E* to a classifier where the ore is separated according to size. Start and stop pushbuttons and a selector switch, which selects one of the two modes of operation, are provided at the master station. A horn warns workers off the conveyors prior to their

Table I. An Operation of the Conveyor

Time, Seconds	Major Events
0......	Set selector switch right, press master start button
1......	Release start button; horn is blowing
5......	Horn stops
10......	Control relay 1013 energized; conveyor *D* starts
15......	Conveyor *C* right drive starts
20......	Crusher starts
25......	Conveyor *A* starts
30......	Conveyor *B* starts
41......	Press lock-out stop button 1006; crusher stops
44......	Conveyor *A* stops
47......	Conveyor *B* stops; *C* and *D* still running
48......	Press master stop button; entire system stops

Table II. Specification Printout

```
CONVEYOR CONTROL SYSTEM        DATA NO. 5                         4 JUNE 1959
TIME, 5 SECONDS              ALL SPECIFICATIONS MET
TIME, 25 SECONDS             ALL SPECIFICATIONS MET
TIME, 45 SECONDS             ALL SPECIFICATIONS MET
    EXCEPT THE FOLLOWING
                             ELEMENT           STATE              SINCE TIME
                               1015              0                    41
TIME, 48 SECONDS             ALL SPECIFICATIONS MET
```

movement, and several emergency lockout stop pushbuttons are installed at various points for safety reasons. The relay switching schematic diagram is shown in Fig. 2. Equipments are interlocked with time-delay relays so that no equipments will start until the downstream equipments are in motion. With the time delays and selector switch positions shown in Fig. 2, the sequence of major events for a possible operation is given in Table I.

PROCEDURES TO USE THE CONTROL PROGRAM

1. Assign numbers to all input elements (e.g., no. 1001 through 1010 of Fig. 1). Assign numbers to all dependent elements (e.g., no. 1011 through 1021 of Fig. 1), and write their transmission functions as illustrated in equations 1 through 11.

$$F_{1011}=[(1001)+(1012)](\overline{1012_0^5}) \tag{1}$$

$$F_{1012}=[(1001)+(1012)](\overline{1012_0^{15}}) \tag{2}$$

$$F_{1013}=(\overline{1002})[(1013)+(\overline{1012_0^{10}})] \tag{3}$$

$$F_{1014}=(1013)[(1017_3^5)+(1018_3^5)](\overline{1004})(1015_3^5) \tag{4}$$

$$F_{1015}=(1013)[(1017_3^5)+(1018_3^5)](\overline{1005})\quad(1016_3^5) \tag{5}$$

$$F_{1016}=(1013)[(1017_3^5)+(1018_3^5)](\overline{1006}) \tag{6}$$

$$F_{1017}=(1013)(\overline{1007})(\overline{1003})(1019_3^5) \tag{7}$$

$$F_{1018}=(1013)(\overline{1007})(1003)(1020_3^5) \tag{8}$$

$$F_{1019}=(1013)(\overline{1003})(\overline{1008}) \tag{9}$$

$$F_{1020}=(1013)(1003)(\overline{1009})(1021_3^5) \tag{10}$$

$$F_{1021}=(1013)(1003)(\overline{1010}) \tag{11}$$

2. Code the input elements and the Boolean functions for the dependent elements in the language of the particular digital computer being used. For the IBM *704* machine, typical examples follow.

For input element 1001:

```
Code                    Comment

PZE 1001,, 2........Two instructions for 1001
TSX INVAR, 4.......Uses input variable subroutine
```

For dependent element 1011 (i.e., equation 1):

```
Code                    Comment

PZE 1011,, 10.......Ten instructions for 1011
TSX DELAY, 4.......Uses delay subroutine
PZE 1012
PZE 5,, 0...........α-delay is 5, β-delay is 0
COM.................Complement the state
STO T...............Save temporarily
CLA 1012
ORA 1001............OR
ANA T...............AND
TRA STATE.........Transfer to control program
```

3. Assemble the code to produce a machine-language program that is compatible with, and will be combined with, the machine-language control program.

4. Punch data cards for the sequential input conditions and specifications for the time duration of the operation. Also enter the total number of elements and the total time of the operation. Provision is made for entering a brief title which may specify the date and conditions of the operation. This description is printed out and also preserved on the magnetic tape that contains the time history of the system.

For the example shown in Fig. 2 and Table I, input conditions were entered at times 0, 1, 41, and 48 seconds. The specifications in Table I at times 5, 25, and 48 seconds were also entered. An incorrect specification (element 1015 is in state *1*) was entered at 45 seconds to illustrate the specification printout from the control program. The printout is shown in Table II.

5. In addition to the specification printout, the control program writes a complete time history of the states of the elements. Although this magnetic tape contains a separate record for each unit of time during the operation, a printout program is provided to enable the engineer to be selective in printing out only those records of special interest as illustrated in Table III. The magnetic tape may be removed from the computer and preserved for further analysis.

Conclusions

1. The control program was tested on the IBM *704* digital computer by simulated operations of switching networks. Assuming an average of ten machine-language instructions for each Boolean function, the present IBM *704* version of the control program determines the states of dependent elements at the rate of about 7,000 per minute. With a 32,768-word IBM *704*, a network of about 3,000 elements can be solved at the high speed.

2. The use of Boolean algebra to describe the transmission of each element in a system provides a better understanding of the relationship of the element to the system. Therefore, the effort to mechanize a system for a simulated operation on a digital computer is well compensated by the advantages of a better designed system.

3. Since time delays are incorporated in the transmission functions written for the control program, most of the relay racing problems in a system can be detected by simulated operation on the digital computer. Also, the required tolerance specifications for the relays can be determined similarly at very low cost because of the high speed operation of the computer.

Table III. Performance Printout

```
TIME 4 SECONDS                 PERFORMANCE FROM TAPE 5
 ELEMENT  STATE  TIME ELEMENT  STATE   TIME  ELEMENT  STATE  TIME   ELEMENT  STATE  TIME   ELEMENT  STATE  TIME
   1001     0      1    1002     0      0      1003     0      0      1004     0      0      1005     0      0
   1006     0      0    1007     0      0      1008     0      0      1009     0      0      1010     0      0
   1011     1      0    1012     1      0      1013     0      0      1014     0      0      1015     0      0
   1016     0      0    1017     0      0      1018     0      0      1019     0      0      1020     0      0
   1021     0      0

TIME 16 SECONDS                PERFORMANCE FROM TAPE 5
 ELEMENT  STATE  TIME ELEMENT  STATE   TIME  ELEMENT  STATE  TIME   ELEMENT  STATE  TIME   ELEMENT  STATE  TIME
   1001     0      1    1002     0      0      1003     0      0      1004     0      0      1005     0      0
   1006     0      0    1007     0      0      1008     0      0      1009     0      0      1010     0      0
   1011     0      5    1012     0     15      1013     1     10      1014     0      0      1015     0      0
   1016     0      0    1017     1     15      1018     0      0      1019     1     10      1020     0      0
   1021     0      0
```

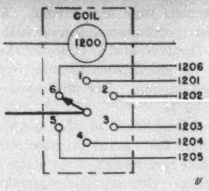


Fig. 3. Six-position stepping switch

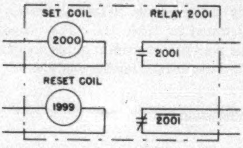


Fig. 4. Mechanical latching relay

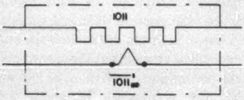


Fig. 5. Detonator or fuse

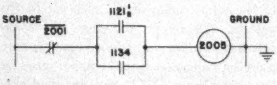


Fig. 6. Permanently grounded relay circuit

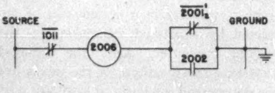


Fig. 7. Ungrounded relay circuit

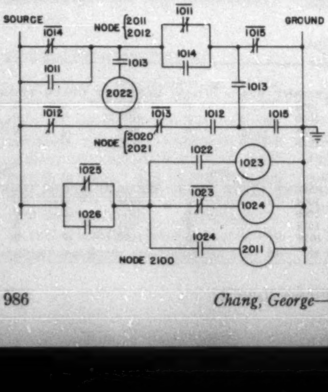


Fig. 8. Relay circuit to illustrate nodal method

4. If adjustable time-delay elements are used in the system, a simulated operation provides the best information for setting the delays for optimum operation of the system.

5. Once the system is coded as a digital program, the performance of the system can be obtained quickly and with very little labor. Therefore, the engineer can afford to analyze the system for all possible input conditions, even those which seldom occur normally.

6. The digital program may be used for systems other than relay and switching if the system can be described by Boolean equations. Therefore, complete systems, including hydraulic and pneumatic equipment, can be tested at high speed and low cost.

Appendix I. Examples of Logical Elements for Common Devices

Stepping Switch

A 6-position rotary-type stepping switch is used for the illustration (see Fig. 3). Each time the coil numbered 1200 receives a pulse, the switch advances its contact one position higher than its previous position. The contacts numbered from 1201 to 1206 are treated as dependent elements of the coil. Their functions are

$$F_{1201}=(1201)(\overline{1200})+(1200)(1206)$$

$$F_{1202}=(1202)(\overline{1200})+(1200)(1201)$$

$$F_{1203}=(1203)(\overline{1200})+(1200)(1202)$$

$$F_{1204}=(1204)(\overline{1200})+(1200)(1203)$$

$$F_{1205}=(1205)(\overline{1200})+(1200)(1204)$$

$$F_{1206}=(1206)(\overline{1200})+(1200)(1205)$$

The first term of the functions indicates that the switch contact remains at its own position, when the coil is not energized. The second term indicates that the switch advances the contact position by one, whenever the coil is energized.

Switching Flip-Flop

A mechanical latching relay is shown symbolically in Fig. 4. The relay numbered 2001 is treated as a dependent element to the set coil 2000 and reset coil 1999. There are two cases possible:

1. Set coil overrides reset coil when both are energized. The function of the relay is

$$F_{2001}=(2001)(\overline{2000})(\overline{1999})+(2000)$$

The Veitch diagram for F_{2001} is

	$F_{1999}=0$	$F_{1999}=1$
$F_{2000}=0$	Unchanged	0
$F_{2000}=1$	1	1

2. Reset coil overrides set coil when both are energized.

$$F_{2001}=\overline{(\overline{2001})(\overline{2000}(\overline{1999})+(1999)}$$

The Veitch diagram for F_{2001} is

	$F_{1999}=0$	$F_{1999}=1$
$F_{2000}=0$	Unchanged	0
$F_{2000}=1$	1	0

Detonator or Fuse

When the heater element 1011 in Fig. 5 is energized, the fuse is assumed to ignite after 1 second. Once the fuse ignites, the circuit in series with the fuse will open indefinitely. With an α-delay of 1 second and β-delay of some time longer than the total simulation time of system operation, this device can be represented in the digital computer program.

Appendix II. Techniques in Writing Transmission Functions

Permanently Grounded Element

The transmission function of the element is the transmission to its source. This theorem is illustrated in Fig. 6 and the transmission function of the relay is

$$F_{2005}=(\overline{2000})[(1121_2{}^1)+(1134)]$$

Ungrounded Element

The transmission function of the element is the AND product of the source transmission and ground transmission. This theorem is illustrated in Fig. 7.

Source transmission $T_s=\overline{1011}$

Ground transmission $T_g=(\overline{2001_2{}^1})+(2002)$

Therefore, the transmission function of the relay is

$$F_{2006}=T_sT_g=(\overline{1011})[(\overline{2001_2{}^1})+(2002)]$$

Nodal Method in Writing Transmission Functions

For a large system, it may be more convenient to write the transmission function of an element from one node to another node instead of tracing the circuit to the source or ground. Since a node can be either connected or disconnected to a source or a ground, the node actually possesses four states. To overcome this difficulty, two numbers may be assigned to a node (i.e., it is resolved into two logical elements). However, if the node cannot be grounded, then one number is sufficient to represent the node. The example in Fig. 8 illustrates an element numbered 2022 in a bridge circuit and the use of the nodal method. Elements 2022, 1023, 1024, and 2011 are voltage coils of their associated relays. Voltage coils are considered to be open-circuit devices, while current coils are short-circuit devices. Transmission functions for the nodes and relays are

$F_{2100}=(\overline{1025})+(1026)$ single number node

$$F_{2011}=[(1011)+(\overline{1014})]+[(\overline{1011})+(1014)](1013)(1012)(\overline{1013})(\overline{1012})$$

But $(1013)(\overline{1013})$ is always zero, then

$F_{2011}=(1011)+(\overline{1014})$; source node

$F_{2012}=[(\overline{1011})+(1014)][(\overline{1015})+(1013)(1015)]$; ground node

$$F_{2020}=(\overline{1012})+(\overline{1013})(1012)(1013)[(\overline{1011})+(1014)][(1011)+(\overline{1014})]$$

$=(\overline{1012})$; source node

$F_{2021}=(\overline{1013})(1012)[(1015)+(1013)(\overline{1015})]$

$=(\overline{1013})(1012)(1015)$; ground node

With the above node transmission functions established, the transmission functions for the relays are extremely simple.

$F_{1023}=(2100)(1022)$

$F_{1024}=(2100)(\overline{1023})$

$F_{2011}=(2100)(1024)$

$F_{2022}=(2011)(2021)(1013)+(2020)(2012)(1013)$

References

1. A SYMBOLIC ANALYSIS OF RELAY AND SWITCHING CIRCUITS, C. E. Shannon. *AIEE Transactions*, vol. 57, 1938, pp. 713–23.

2. SWITCHING CIRCUITS AND LOGICAL DESIGN (book), S. H. Caldwell. John Wiley & Sons, Inc., New York, N. Y., 1958.

3. LOGICAL DESIGN OF DIGITAL COMPUTERS (book), M. Phister. John Wiley & Sons, Inc., 1958.

4. DIGITAL COMPUTER PROGRAMING (book), D. D. McCrckean. John Wiley & Sons, Inc., 1957.

Transistorized Multifrequency Ringing Generator

J. F. KOSTELICH
ASSOCIATE MEMBER AIEE

B. W. HOWALD
ASSOCIATE MEMBER AIEE

THE selective telephone party-line ringing system using mechanically tuned bells is relatively well known. A 5-party line, for example, may have phones with bells tuned to 20, 30, 40, 50, and 60-cycles, respectively. Introducing a 40-cycle ringing voltage to this party line will cause only the 40-cycle bell to ring.

Several types of ringing voltage generators may be used to supply the required frequencies for a telephone central office. Both a-c and d-c driven motor-generator sets, vibrator-type inverters, magnetic frequency changers, and electronic-tube-type generators are in use today. The reasons for developing still another type of ringing voltage generator can be best explained by outlining the requirements of the ideal generator, together with the basis for each requirement.

Some of these requirements are as follows:

1. The output frequency of the generator should be stable and capable of precise adjustment to the required frequency. The sharp mechanical and electrical resonance curves of the selective ringer dictates that the ringing frequency should not shift more than 1/2 cycle from its initial value under any normal operating condition. Preferably, the shift should be kept within ±1/4 cycle with varying load, ambient temperature, input voltage, and age. In addition, the frequency setting must be precise because in most installations two generators will be installed, one for normal and one for stand-by use, and the performance of the system should not change when transferring from normal to stand-by operation. The normal and stand-by generators must thus be adjusted to the same frequency.

2. The output voltage of the generator should be nearly free of harmonics. There are two reasons for this requirement: the elimination of cross-ring and the elimination of noise and crosstalk in the telephone system.

Three series of ringing frequencies are now in common use: the harmonic series, using $16^2/_3$, 25, $33^1/_3$, 50, and $66^2/_3$ cycles; the synchromonic series, using 20 (or 16), 30, 42, 54, and 66 cycles; and the decimonic series, using 20, 30, 40, 50, and 60 cycles. In two of these three series the second and third harmonics of the lower frequencies can cross-ring the bells of the higher frequencies. The amount of voltage required to produce an objectionable cross-ring, or tinkle of the ringer, which should remain silent, is less than 10 volts with some ringers. Since the ringing voltages of the lower frequencies are about 100 volts, the second and third harmonics in the lower frequencies must be kept well below 10%.

Fig. 1 shows the selectivity curves of typical telephone ringers. With 60 volts applied, the 20-cycle and 40-cycle ringers are frequency selective within 3 cycles. Although the bell selectivity curves appear to be broad, it must be kept in mind that on occasion bells are tuned in the field, and the setting selected may be just inside the selectivity curve limits. Thus the ringing generator frequency must be stable, as a significant frequency shift in ringing voltage may cause a ringing failure in these field-tuned bells. Note that 15 volts of second harmonic voltage in the 20-cycle ringing voltage would be sufficient to energize the 40-cycle ringer.

The harmonics in the higher audio range are also objectionable as they produce noise

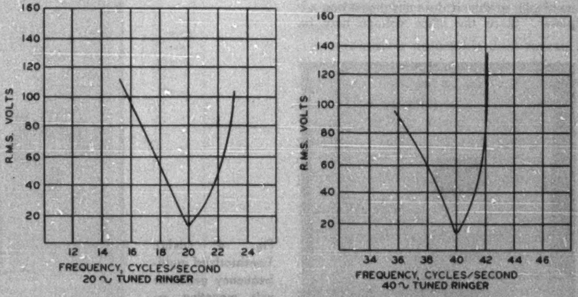


Fig. 1. Frequency selectivity curves for typical tuned ringers

---

Paper 59-1175, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted June 9, 1959; made available for printing August 17, 1959.

J. F. KOSTELICH and B. W. HOWALD are with Lorain Products Corporation, Lorain, Ohio.

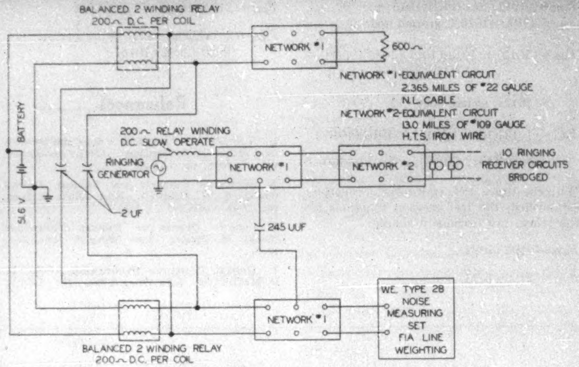


Fig. 2. Crosstalk noise-measurement test circuit

or crosstalk in the adjacent pairs in the telephone cable. The Rural Electrification Administration (REA) requirement is that the noise measured on an equivalent cable circuit, as shown in Fig. 2, must not exceed 25 dba (decibels adjusted) as measured with the Western Electric *2B* noise measuring set (*F1A* line weighting).

3. The generator should be dependable and not require frequent maintenance or adjustment. Ideally, it should not have any moving or expendable parts, and all components should have anticipated life of the order of 50,000 hours or more. If maintenance is required, it should be simple in nature, so that a unit which is out of service can be replaced or repaired in a very few minutes.

4. The generator should be self-protecting in case of overloads or short circuits so that it will not be damaged by ringing into short-circuited lines or by faulty interrupters or accidental short circuits placed across it during maintenance operations in the office. Preferably, the protection should not be in the nature of fusing as the replacement of fuses in an unattended office can be a serious nuisance.

5. The generator preferably should operate on the telephone office battery (50 volts d-c nominal), so that it does not depend on a-c power. Since the office voltage may be anywhere between 44 and 56 volts, the generator must operate satisfactorily and not be damaged by operation over this range of voltage.

6. The generator should have good voltage regulation so that satisfactory ringing will be obtained under all load conditions. On long, heavily loaded lines, where not all the ringers are of the same type, it is sometimes difficult to get satisfactory ringing on all phones. In these cases, well-regulated ringing voltage is important, so that ringer adjustments, made under one load condition, are still correct when the load is considerably different, or when the office battery voltage drops during an a-c power failure or rises while the batteries are being equalized. For ideal operation, the ringing voltage should not change with the load or with changes in input voltage.

7. The power consumption of the ringing generator should be as small as possible, especially if it operates from the office battery. The no-load drain is important as the generator may operate for considerable periods with little or no load. Since start-stop operation is frequently employed, the voltage and frequency should not shift appreciably during the warm-up period.

It is the purpose of this paper to describe a transistorized multifrequency ringing generator, which may be the proper approach to this idealized source of ringing voltage.

Fig. 3 shows the transistorized ringing generator to be described. This unit mounts on a 19-inch relay rack, requires 17½ inches of vertical space, and is 13⅝ inches deep. Fig. 4 shows the slide mounting arrangement of the units. The five frequency generators are identical and interchangeable, while the sixth unit, in the lower right-hand corner, contains the battery filter, a reverting tone generator, and remote control features.

Fig. 5 is a block diagram of the 5-frequency ringing machine. Power input into the generators is supplied through a battery filter and remote start relay in the common unit. The remote start relay can be energized remotely by applying (positive) ground to the start terminal. The identical ringing generators are shown with their output ringing voltage appearing on the terminal block as indicated.

The output of the audible ringback tone generator appears on the terminal block. The tone generator output is 420 cycles with 40-cycle amplitude modulation, which is the industry standard for reverting tone.

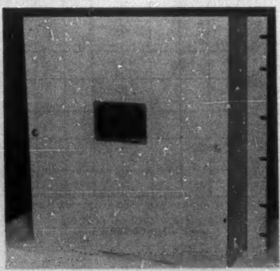


Fig. 3 (left). Transistorized multifrequency ringing generator

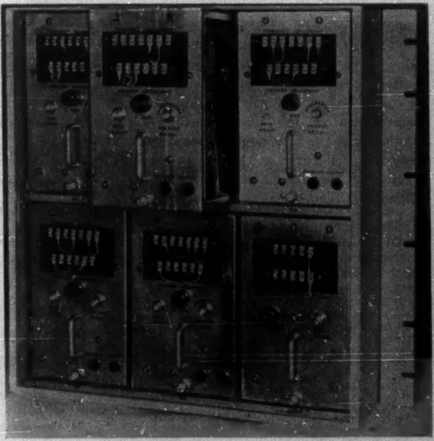


Fig. 4 (right). Transistorized multifrequency generator unit mounting arrangement

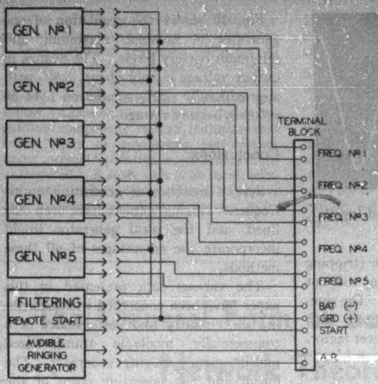

GEN. Nº1
GEN. Nº2
GEN. Nº3
GEN. Nº4
GEN. Nº5
FILTERING
REMOTE START
AUDIBLE RINGING GENERATOR
TERMINAL BLOCK
FREQ. Nº1
FREQ. Nº2
FREQ. Nº3
FREQ. Nº4
FREQ. Nº5
BAT (−)
GRD (+)
START
A.R.

Fig. 5. Block diagram of multifrequency generator

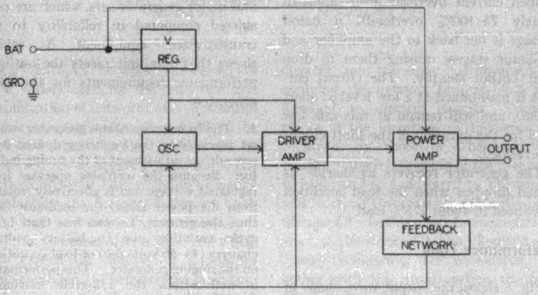

BAT
GRD
V REG.
OSC
DRIVER AMP
POWER AMP
OUTPUT
FEEDBACK NETWORK

Fig. 6. Block diagram of a single-frequency generator unit

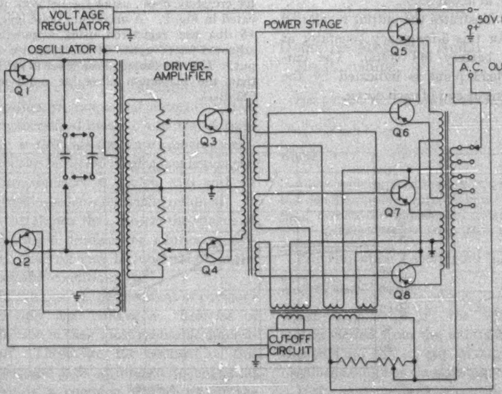

VOLTAGE REGULATOR
OSCILLATOR
DRIVER-AMPLIFIER
POWER STAGE
50V. D.C.
A.C. OUTPUT
Q1
Q2
Q3
Q4
Q5
Q6
Q7
Q8
CUT-OFF CIRCUIT

Fig. 7. Schematic diagram of a single-frequency generator unit

Circuit Performance

A block diagram of a transistorized frequency generator circuit is shown in Fig. 6 while Fig. 7 shows a partial schematic diagram of the generator unit.

Voltage Regulator

A simple voltage regulator, comprised of a Zener diode and a power transistor, provides a constant 12-volt direct current to the oscillator and driver-amplifier stages. The regulator is required to keep the output alternating voltages of the oscillator and driver stages constant during battery voltage variations (44 to 56 volts d-c).

Oscillator

The oscillator is a tuned collector oscillator. The frequency is determined by an *L-C* tank circuit, which can be adjusted readily from 15 to 66²/₃ inches. Coarse tuning is accomplished by capacitor tap selection. The two capacitors in the circuit are used singly, or they are used in parallel. The fine tuning is done by a screwdriver adjustment on the variable gap inductor. The a-c output voltage of the oscillator is constant at all frequencies to which the oscillator is tuned.

Two 250-milliwatt transistors work in push-pull, with a common base configuration being used to minimize the effect of transistor parameter variations.

Driver-Amplifier

The driver-amplifier stage is transformer-coupled to the oscillator. This stage is a common-collector current-amplifier circuit. The regulation of this stage must be good, so that the drive voltage supplied to the next stage will not decrease with increased load requirements. This high-gain stage requires little power from the oscillator, so that the oscillator is effectively isolated from the power amplifier by the driver amplifier stage.

A screwdriver adjustable dual resistor varies the input drive to the driver-amplifier. This is a fine adjustment providing approximately 15% adjust variation for the output voltage of the generator. Included in this stage is part of the cutoff circuitry for short-circuit protection of the generator as a whole.

Power Amplifier

The power stage, transformer-coupled to the driver-amplifier stage, is comprised of a 4-transistor bridge, output transformer, and current transformer.

The power transistors operate class *B*,

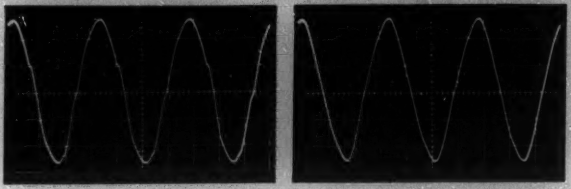


Fig. 8. Output waveshape, 50 cycles at 125 volts full load (left) and no load (right)

producing a sine-wave output voltage. Each transistor has a separate drive winding on the driver amplifier output transformer. Inverse feedback windings on the power transformer, in conjunction with the drive windings, provide a constant drive voltage on the power transistors. To compensate for circuit losses and decreasing current gain in the power transistors at higher generator load levels, a positive current feedback winding is also used in the power transistor drive circuit.

The adjustable positive current feedback feature allows setting the output voltage at nearly flat regulation from no load to full load on the generator.

The constant drive voltage results in good output voltage regulation with wide variations in input battery voltage. With a 20% change in battery voltage (44–56 volts d-c), the change in a-c output voltage is about ±5%.

To promote maximum efficiency in the power stage, the drive voltage is set at a maximum value that will allow a sine-wave output and yet not produce clipping of the output wave at low battery levels or high load conditions. In this way transistor heat dissipation is kept to a minimum. Adequate heat sink area has been used to maintain transistor junction temperature to a safe value at all ambient and load conditions.

The output transformer design is such that pulses of direct current (tripping battery) in the output winding will not distort the output waveshape. The sine-wave output is also maintained with reactive loads down to 70% power factor.

CUTOFF CIRCUIT

The cutoff circuit in this generator is a current-sensitive device. At a prescribed current overload level (approximately 75–100% overload), a cutoff voltage is fed back to the amplifier and oscillator stages, causing them to drop their output sharply. The output current is maintained at a low level at short circuit, and will remain at this safe low level for the duration of the short circuit or overload.

The generator recovers as sharply as it had dropped when the load condition is reduced to about 175% load.

Performance Data

Fig. 8 shows the output wave shape of a frequency generator set at 125 volts, 50 cycles, at no load and full load. A reasonably good sine wave is produced, whose third harmonic content averages very low, at about 5.2%.

Fig. 9 illustrates the output regulation curves for this 5-frequency generator at nominal battery (48 volts). The cutoff overload point is indicated by the dotted line at end of each curve.

Fig. 10 shows the regulation of one generator at nominal, maximum, and minimum battery levels. The change in output voltage is only about ±5% with approximately 20% variation (44–56 volts) in battery voltage.

Conclusions

Several methods of generating multifrequency ringing voltages have been outlined, and the ideal generator would incorporate the advantages of all these methods.

The various data presented in this paper have been concerned with describing the circuitry and performance of a commercially available transistorized multifrequency ringing generator. The initial and the operating costs of the equipment described are generally less than those of magnetic frequency changers and motor-generator sets, which are considered compared in reliability to the transistorized equipment. A review shows that this unit meets the outlined performance requirements for the ideal generator:

1. The frequencies of this generator can be set precisely to the frequency desired by a screwdriver adjustment of the tuning inductor. Because the oscillator operates from regulated voltage and is effectively isolated from the power stage, the oscillator (and thus the generator) shows less than 1/10-cycle variation due to battery voltage changes (44–56 volts d-c) or load conditions on the ringing generator. This performance is well within the 1/2-cycle maximum frequency shift requirement.

2. The output voltage is a sine wave. The harmonic content averages 5.2%, which is not enough to cause cross-ring difficulty. Laboratory checks were made for crosstalk noise, using the network indicated in Fig. 2. A maximum noise level of 15 dba was registered using a generator adjusted for 66 cycles at 140 volts a-c output. This maximum noise level is far less than the maximum allowable noise level (25 dba).

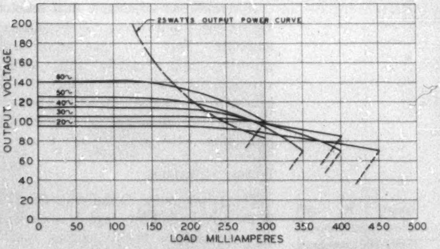


Fig. 9. Output voltage regulation curves for multifrequency generator

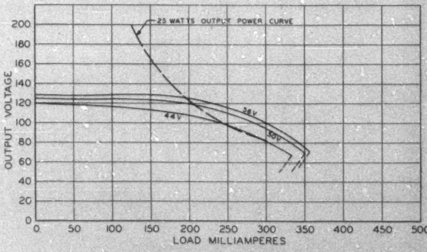


Fig. 10. Output voltage regulation curve for 50-cycle generator with different input battery voltages

3. One of the original multifrequency generators has been installed on field trial in a local 4,000-line telephone office for 20 months. To date, there has been no opertional difficulty or need for maintenance reported by the telephone operating personnel. Conservative electrical design regarding component ratings indicates that many years of troublefree performance under normal operating conditions can be anticipated. A spare unit is on hand, however, and can be quickly adjusted to replace any frequency generator if service is required.

4. The generator is self-protecting in case of overloads or short circuits on the generator output.

5. The generator operates from telephone office battery (50 volts d-c nominal), eliminating its dependence on a-c power.

6. The voltage regulation of the generator can be adjusted to less than a 5-volt variation in output voltage from no-load to full-load conditions. Changes in battery voltage (44–56 volts d-c) have a small effect on output voltage regulation, as shown in Fig. 10.

7. The power consumption per frequency generator is 100 milliamperes at no load, and about 1 ampere at full load (25 watts). The over-all efficiency of the multifrequency generator is about 55% at full load.

Input battery filtering is provided in the common unit. With full load on the multifrequency generator, battery noise of 18 dba maximum was read with a Western Electric *2B* noise measuring set (*F1A* line weighting). The filtering is adequate to keep the battery noise level well below the maximum level of 32 dba.

The General Form of Some Common Network Theorems

COLIN ADAMSON
MEMBER AIEE

S. M. EL-SOBKI
NONMEMBER AIEE

WHEREAS network theorems are widely used for the reduction and simplification of networks of a substantial degree of complexity, they tend to be neglected when complex systems are to be solved on a network analyzer. The network analyzer has proved to be a very satisfactory tool for power network analysis, but a review of the situation is profitable for cases of large problems which may exceed the number of impedance units available. There are two main general aspects to the use of network theorems:

1. The use of theorems for the specific investigation of a portion of a network, or the relationship between two portions of a network, e.g., Thévenin's, or the compensation, theorem.
2. The use of techniques of successive reduction, e.g., mesh-to-star transformation.

The most significant application of the reduction techniques of category (2) is the generalized mesh-to-star transformation; in this case there are numerous conditions to be satisfied before reduction can be accomplished if a plain star without mutual impedances is finally required. It can be shown that during the stages of preparing a problem for solution, a network analyzer may be used to obtain the necessary conditions.

In the case of the theorems of category (1), they also behave as theorems of reduction in their most profitable applications. These are the concern of this paper, and it is of interest to note that there is a common thread to all the theorems in this category, and that they, and their derivatives, may all be deduced directly from the equations obtained by applying the network laws of Ohm and Kirchoff to electric circuits. Premier places in the hierarchy of theroems are occupied by those of Thévenin and Norton;[1] the relationship of the theorems of Millman,[2] and of Starr,[3] to Thévenin's and Norton's theorems have recently been discussed.[4] In this paper further relationships to the work of Ataka,[5] Zepler,[6] and Tasny-Tschiassny,[7] are considered, together with the compensation theorem. The common threads of all the network theorems in category 1 are thus exposed, together with the methods to be adopted to obtain network reduction in a practical manner; illustrative practical examples are quoted in the Appendix.

Definitive Mesh and Nodal Relationships

For a mesh network with *m* meshes:

$$\|e_m\| = \|z_{mm}\| \times \|i_m\| \quad (1)$$

where

$\|e_m\|$ denotes impressed mesh voltages.
$\|i_m\|$ denotes response currents in the meshes.
$\|z_{mm}\|$ defines the impedance matrix of the network viewed from its *m* meshes.

From equation 1 solving for $\|i_m\|$ yields

$$\|i_m\| = \|z_{mm}\|^{-1} \times \|e_m\| \quad 1(A)$$

where $\|z_{mm}\|^{-1} = \|a_{mm}\|$, the admittance matrix viewed from the *m* meshes. The elements $a_{11}, a_{21} \ldots$, etc., of a_{mm} may be obtained readily on the network analyzer.[8]

If the same network is considered, the number of node pairs being *n*, the performance equation of the network is

$$\|I_n\| = \|Y_{nn}\| \times \|V_n\| \quad (2)$$

where

$\|I_n\|$ denotes current through the *n* node pairs.
$\|V_n\|$ denotes voltage across the pairs of nodes.
$\|Y_{nn}\|$ is the admittance matrix viewed from the *n* node pairs.

The practical method of obtaining the elements Y_{11}, Y_{21}, etc, of $\|Y_{nn}\|$ is demonstrated later.

Thévenin's Theorem

The generalization of Thévenin theorem has been discussed recently;[4] the main points of interest in this discussion were the relationship of the theorem with the technique of "tearing" advocated by Kron,[9] and the inclusion in the general body of the theorem of the special cases of Millman and Starr. Thévenin's theorem will be treated here briefly by matrix methods to illustrate, first, its relationship to the work of Ataka, Tasny-Tschiassny, and Zepler, and to the compensation theorem.

Fig. 1 shows a general network having *m* meshes, with mesh currents i_{m1}, i_{m2}, etc., and *n* accessible node pairs across which there are voltages $V_1, V_2, \ldots, V_q\ V_r, V_k$, etc. If the accessible node pairs are traced through to an external circuit, lying to the right of, but not shown in, Fig. 1 and connected to the same *n* node pairs, then *n* meshes are added to the original networks; this is equivalent to considering the network of Fig. 1 as containing $(m+n)$ meshes. It should be noted that the additional of the external circuit to give an additional *n* meshes,

---

Paper 59-1180, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted July 24, 1958; made available for printing August 11, 1959.

COLIN ADAMSON and S. M. EL-SOBKI are with Manchester College of Science and Technology, Manchester, England.

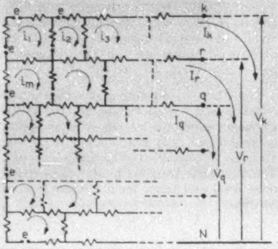


Fig. 1. General network with m meshes and n node pairs

makes the n mesh currents the same as the n currents flowing in the n independent node pairs; thus $i_n = I_n$. The technique may be better appreciated by reference to Fig. 2, where e_k' is the sum of all sources of emf (electromotive force) encountered within the mesh k; here the kth mesh containing terminal k has been considered and the cumulative mesh emf has been obtained by traversing the mesh in a positive direction and summating all the emf including the terminal voltage. Thus $e_k = e_k' - V_k$.

Obviously, the relationship between mesh voltages (of the $m+n$ meshes), the matrix impedance of the complete network, and the mesh currents is

$$\|e\| = \|z\| \times \|i\| \tag{3}$$

Equation 3 may be partitioned along the m and n meshes, the latter being those which contain the accessible terminals:

$$\begin{Vmatrix} e_m \\ e_n \end{Vmatrix} = \begin{Vmatrix} z_{mm} & z_{mn} \\ z_{nm} & {}_{n\ n} \end{Vmatrix} \times \begin{Vmatrix} i_m \\ I_n \end{Vmatrix} \tag{3(A)}$$

or

$$\|e_m\| = \|z_{mm}\| \times \|i_m\| + \|z_{mn}\| \times \|I_n\| \tag{3(B)}$$
$$\|e_n\| = \|z_{nm}\| \times \|i_m\| + \|z_{nn}\| \times \|I_n\|$$

where

$$\|e_n\| = \|e_n'\| - \|V_n\| \tag{3(C)}$$

and represents the total of equations such as the one, $e_k = e_k' - V_k$ obtained from the kth node pair.

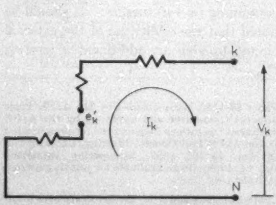


Fig. 2. The kth mesh of the general network

There are two cases of major interest when these equations are solved:

1. The relationship between $\|I_n\|$ and $\|V_n\|$, i.e., the conditions at the terminals; this is Thévenin's theorem.
2. The relationship between $\|i_m\|$ and $\|V_n\|$ i.e., the effect of terminal conditions on the network, which is useful when the effect of connecting loads on load flow and losses are being considered.

For case (1), eliminating $\|i_m\|$ yields

$$\|V_n\| = \|e_n'\| - \|z_{nm}\| \times \|z_{mm}\|^{-1} \times \|e_m\| - \{\|z_{nn}\| - \|z_{nm}\| \times \|z_{mm}\|^{-1} \times \|z_{mn}\|\} \times \|I_n\| \tag{4}$$

Equating $\|I_n\|$ to zero to give the open-circuit conditions at the nodes yields

$$\|V_n\|_{oc} = \|e_n'\| - \|z_{nm}\| \times \|z_{mm}\|^{-1} \times \|e_m\| \tag{5}$$

Equating all the mesh emfs to zero, i.e.,

$\|e_n'\| = \|e_m\| = 0$, yields

$$\|V_n\| \times \|I_n\|^{-1} = \|z_{nn}\| - \|z_{nm}\| \times \|z_{mm}\|^{-1} \times \|z_{mn}\| = \|Z_{nn}\| \tag{6}$$

where $\|Z_{nn}\|$ represents the impedance of the network viewed from its n node pairs. Therefore

$$\|V_n\| = \|V_n\|_{oc} - \|Z_{nn}\| \times \|I_n\|$$

Also, $\|V_n\| = \|Z_{nn}'\| \times \|I_n\|$ where $\|Z_{nn}'\|$ is the load impedance. Therefore

$$\|V_n\| = \{\|Z_{nn}'\| + \|Z_{nn}\|\} \times \|I_n\| \tag{7}$$

which is a general statement of Thévenin's theorem.

For case (2), the relationship between $\|i_m\|$ and $\|V_n\|$ may be obtained as follows. By applying voltages $\|e_n''\|$, each of one pu (per unit) value, between each node and the reference node, equations 3(B) become

$$\|e_m\| = \|z_{mm}\| \times \|i_m\| + \|z_{mn}\| \times \|I_n\|$$
$$\|e_n'\| - (\|V_n\| + \|e_n''\|) = \|z_{nm}\| \times \|i_m\| + \|z_{nn}\| \times \|I_n\|$$

By equating the original sources of emf in the network, $\|e_m\|$ and $\|e_n'\|$, to zero, and solving for $\|i_m\|$, the following expression is obtained:

$$-[\|z_{mm}\| - \|z_{mn}\| \times (\|Z_{nn}'\| + \|z_{nn}\|)^{-1} \times \|z_{nm}\|]^{-1} \times \|z_{mm}\| \times (\|Z_{nn}'\| + \|z_{nn}\|)^{-1} \times \|e_n''\| = \|i_m\| \tag{8}$$

The mesh circulating currents $\|i_m\|$ represent the transfer admittances in pu between the n node-pairs and the m meshes.

Also, rewriting equation 3(A) yields

$$\|e_m\| = \|z_{mm}\| \times \|i_m\| + \|z_{mn}\| \times \|I_n\|$$
$$\|e_n'\| = \|z_{nm}\| \times \|i_m\| + (\|Z_{nn}'\| + \|z_{nn}\|) \times \|I_n\| \tag{9}$$

Eliminating $\|I_n\|$ and solving for $\|i_m\|$ in terms of $\|V_n\|_{oo}$ and $\|i_m\|_o$

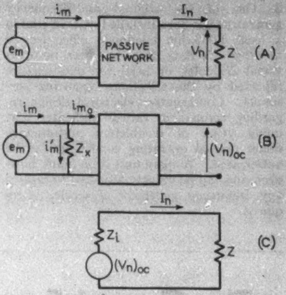


Fig. 3. A—Linear network with m=1 and n=1. B—Zepler's equivalent representation of (A). C—Thévenin equivalent of (A) and (B)

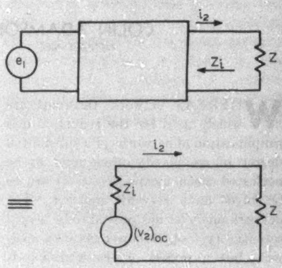


Fig. 4. Equivalent circuit for obtaining Y_t

which is the current in the meshes when $\|I_n\| = 0$, yields

$$\|i_m\| = \|i_m\|_o + \|Y_t\| \times \|V_n\|_{oc} \tag{10}$$

where $\|Y_t\|$ is the transfer admittance between the n node-pairs and m meshes as given in equation 8; then the change in mesh current (or branch current) due to load connection is given as follows: $\|Y_t\| \times \|Y_n\|_{oc}$.

Similar reasoning gives the dual of this relationship: the nodal voltage due to load connection at the nodes is

$$\|V_m\| = \|V_m\|_{sc} + \|Z_t\| \times \|I_n\|_{sc} \tag{11}$$

where $|V_m|_{sc}$ is the voltage at the m node pairs when the n node pairs are short-circuited, and where $|Z_t|$ is the mutual impedance between the m node pairs and the n node pairs, and $\|I_n\|_{sc}$ is the short-circuit current at the n node pairs.

Some General Cases

The foregoing derivations include many special cases which have received special formulation; two of these, the theorems of Millman and Starr, have been discussed

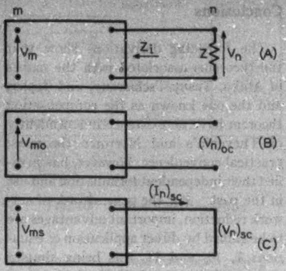


Fig. 5. A—Linear active netork with m=1 and n=1. B—Network of (A) with open-circuit conditions at node pair n. C—Network of (A) with short-circuit conditions at node-pair n

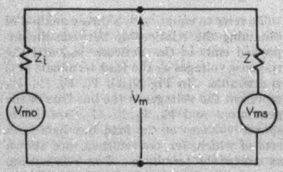


Fig. 6. Tasny-Tschiassny's equivalent circuit of Fig. 5(A) for obtaining V_m

recently, and shown to have no existence independently either of Thévenin's theorem or steps leading to the formulation of Thévenin's theorem: thus, for example, equation 7 for the case of $n=3$ gives Starr's theorem. The four other cases of note are discussed in the following sections.

ZEPLER'S THEOREM

Zepler's network theorem, formulated in 1952, evaluated for a generator connected between two nodes, the effect of connecting a static load between two other nodes. This is thus a special case of Thévenin's theorem but with a limitation in that it is inapplicable to active networks.

It states that in a linear network containing a generator e_m, an impedance Z may be replaced, without changing the effect on the generator, by another impedance Z_z connected in parallel with the generator, where

$$Z_z=\left(\frac{e_m}{(V_n)_{oc}}\right)^2\times(Z_t+Z) \tag{12}$$

where Z_t and $(V_n)_{oc}$ are the two quantities which specify the Thévenin equivalent circuit viewed from the load side; Fig. 3.

Writing $\|i_m'\|$ for the product $\|Y_t\|\times\|V_n\|_{oc}$ in equation 10 yields

$$\|i_m\|=\|i_m\|_0+\|i_m'\| \tag{10A}$$

where $\|i_m'\|$ are the currents resulting from the connection of impedances $\|Z_{nn}\|$.

Applying equation 10(A) to Fig. 3(A) (a circuit with only one pair of nodes and one mesh of interest) yields

$$i_m=i_{m0}+i_m' \tag{10B}$$

where i_m' is the current in the equivalent network, Fig. 3(B), which results from the connection of the load impedance Z. To obtain i_m', Y_t must be obtained, as follows.

Referring to Fig. 4,

$$Y_t=\frac{i_2}{e_1}=\frac{(v_2)_{oc}}{(Z_t+Z)}\times\frac{1}{e_1} \tag{13}$$

Thus

$$i_m'=\frac{(v_2)_{oc}}{e_1(Z_t+Z)}\times(V_n)_{oc} \tag{14}$$

If the network is passive and has only one source of emf, then $e_1=e_m$; $(v_2)_{oc}=(V_n)_{oc}$, and equation 14 becomes

$$i_m'=\frac{(V_n)_{oc}^2}{e_m(Z_t+Z)}$$

$$=\left(\frac{(V_n)_{oc}}{e_m}\right)^2\times\frac{e_m}{Z_t+Z}=\frac{e_m}{Z_z}$$

where Z_z is the impedance given by Zepler's theorem. It can thus be seen that the theorem is a special case of equation 10.

It is interesting to note that there is no general formulation of Zepler's theorem for cases of more than one pair of nodes and more than one internal mesh of salient interest. An equivalent system may, however, be created which follows directly from equation 10, the load impedance being replaced in the meshes which are of interest by currents injected from external current sources.

TASNY-TSCHIASSNY'S THEOREM

The Tasny-Tschiassny theorem has been regarded by its originator as a useful extension of Thévenin theorem and is as follows; see Figs. 5 and 6:

> If an impedance Z is connected between two terminals of a network, the potential difference between any second pair of terminals will be equal to that between the common terminals of two generators connected in parallel and having emfs equal to the voltages across the second pair of terminals when $Z=\infty$ and when $Z=0$ and internal impedance equal to the Thévenin impedance, Z_t, of the network viewed from the first pair of terminals and to Z, respectively.

Thus, from Figs. 5 and 6

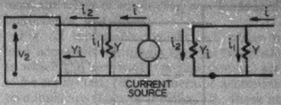


Fig. 7. Relation between Y_t and Y_t'

$$V_m=V_{ms}+(V_{m0}-V_{ms})\times\frac{Z}{Z_t+Z} \tag{15}$$

$$=V_{m0}+(V_{ms}-V_{m0})\times\frac{Z_t}{Z_t+Z}$$

These relationships may be derived from equation 11 for the case when $m=1$ and $n=1$, i.e.,

$$V_m=V_{ms}+Z_t\times(I_n)_{SC} \tag{11A}$$

Referring to Fig. 7 and applying a current source across the pair of nodes n, and measuring the voltage v_2 at the other node-pair m yields

$$Z_t=\frac{v_2}{i}\text{ and }Z_t'=\frac{v_2}{i_2}$$

where Z_t' is the mutual impedance when the load impedance $Z=\infty$, i.e.,

$$Z_t'=\frac{v_2}{i}\times\frac{Y_t+Y}{Y_t}$$

$$=Z_t\times\left(\frac{Z+Z_t}{Z}\right)$$

Applying equation 11(A), when $Z=\infty$, yields

$$V_{m0}=V_{ms}+Z_t'\times(I_n)_{SC}$$

$$V_{m0}=V_{ms}+Z_t\times\frac{(Z+Z_t)}{Z}\times(I_n)_{SC}$$

Therefore

$$Z_t\times(I_n)_{SC}=(V_{m0}-V_{ms})\times\frac{Z}{(Z+Z_t)}$$

By substituting in equation 11(A) the Tasny-Tschiassny equations can be obtained, i.e.,

$$V_m=V_{ms}+(V_{m0}-V_{ms})\times\frac{Z}{(Z+Z_t)}$$

and

$$V_m=V_{m0}+(V_{ms}-V_{m0})\times\frac{Z_t}{(Z+Z_t)}$$

and an alternative form

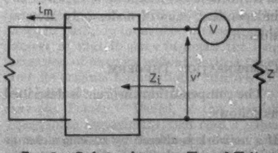


Fig. 8. Relation between Z_t and Z_t''

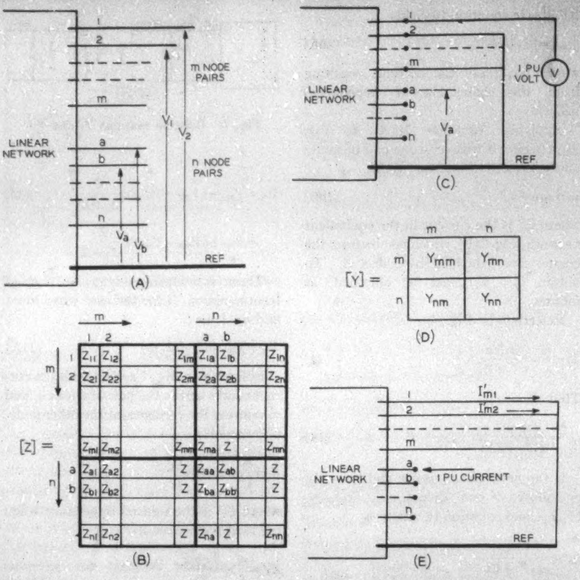


Fig. 9. A—Linear network with m generator node-pairs, and n load node-pairs. B—Matrix impedance of (A). C—Arrangements for obtaining the voltage distribution factor for (A). D—Matrix admittance of (A). E—Arrangements for obtaining the current distribution factor of (E)

$$V_m = V_{m0} - (V_n)_{\mathrm{oc}} \times \frac{Z_t'}{Z+Z_t}$$

Thus this theorem is a special case of equation 11, for $m=1$ and $n=1$, i.e., two independent pairs of nodes only.

ATAKA'S THEOREM

Ataka's theorem has also been formulated as an extension of Thévenin's theorem, as follows:[5]

> A current i_m through any branch m, say, is equal to the sum of a current i_{m0} through the branch m, when any other branch n, say, is open-circuited and a current i_m' through the branch m when an open-circuited emf V_{n0} acts in the branch n and all other emfs are replaced by their internal impedances.

$$i_m = i_{m0} + \frac{(V_n)_{\mathrm{oc}}}{Z_t}$$

It can be readily seen that this theorem is a special case of equation 10 when two independent pairs of nodes only are considered.

COMPENSATION THEOREM

The compensation theorem is described as follows.

> If a network is altered by making a change ΔZ in the impedance Z of a branch, the change in the currents in all other branches is that which woud be produced by an emf of value $-i \times \Delta Z$ acting in series with the modified branch, where i is the current in Z.

As the theorem stands in this form it cannot represent the effect of changing Z from ∞ to Z but, with the use of equation 10, the change in the current in any branch m, say, due to change of impedance of another branch n, say, from Z_1 to Z_2 is given by

$$I_{m1} - I_{m2} = \frac{Z_t}{Z_t''} \times (V_n)_{\mathrm{oc}} \times \left(\frac{1}{Z_t+Z_1} - \frac{1}{Z_t+Z_2}\right) \quad (10B)$$

This equation covers the whole range of change of Z from 0 to $\pm\infty$; Z_t'' represents the transfer impedance between branch m and branch n, when Z (of branch n) $=0$.

To relate Z_t'' to Z_t, consider Fig. 8:

$$Z_t = \frac{v}{i_m}$$

$$Z_t'' = \frac{v'}{i_m} = \frac{v}{i_m}\left(\frac{Z_t}{Z_t+Z}\right)$$

$$Z_t'' = Z_t \times \frac{Z_t}{(Z_t+Z)}$$

Conclusions

The foregoing derivations show that the theorems associated with the names of Ataka, Tasny-Tschiassny, and Zepler, and the one known as the compensation theorem have no existence independently of Thévenin's and Norton's theorems; practical convenience, however, has justified their independent formulation and use in the past. For the general task of network reduction, important advantages are to be gained by direct application of equations 7, 10, and 11, this being simpler than applying the individual theorems independently.

Appendix

Voltage Distribution Factor

To refer to equation 4, a direct method of obtaining the relationship between the impressed emfs in the network $|e_m|$ and the response voltages at the load terminals $|V_n|$ is desirable. In Fig. 9(A), V_1, V_2, ... V_m represent the voltages at the bus bars of the generators, and V_a, V_b, ... V_n, are the response voltages at the load bus bars both sets of which, for convenience, are shown as accessible terminals. The relationship between the two sets of voltages may be obtained as follows.

By impressing 1 pu current at each accessible generator terminal in turn, the values of the response voltages across the various pairs of load terminals will give the pu values of the elements of $|Z_{nm}|$, where this impedance is the bottom left-hand section of Fig. 9(B), i.e.,

$$\|V_n\| = \|Z_{nm}\| \times \|I_m\|$$

but

$$\|I_m\| = \|Z_{mm}\|^{-1} \times \|V_m\|$$

where $|Z_{mm}|$ is obtained in the same manner, and is the upper left-hand section of Fig. 9(B). Thus

$$\|V_n\| = \|Z_{nm}\| \times \|Z_{mm}\|^{-1} \times \|V_m\|$$

The relationship between $|V_m|$ and $|V_n|$, i.e., $|Z_{nm}||Z_{mm}|^{-1}$ is thus obtained directly, as shown in Fig. 9(C), by short-circuiting $(m-1)$ generator nodes to neutral in turn and measuring the voltages $|V_n|$ with 1-pu voltage applied between the generator node and neutral. In the event of loads being connected to the generator bus bars, this procedure still holds, provided that the loads are correctly represented at terminals 1, 2, etc., but are still contained within the "box" shown in Fig. 9(C).

Current Distribution Factor

In the case of the current distribution factor, the relationship between the load current $|I_n|$ and the currents at the generator terminals $|I_m'|$, due to the flow of current $|I_n|$, can be evaluated as follows: By impressing 1-pu voltage between the load terminals n and the neutral, in turn (all other terminals being short-circuited), and

Vol.
Pt. 1
195

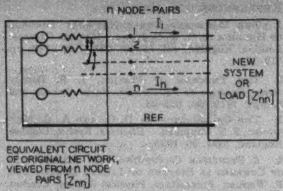

n NODE-PAIRS
I₁
1
2
Iₙ
n
NEW SYSTEM OR LOAD [Z'ₙₙ]
REF
EQUIVALENT CIRCUIT OF ORIGINAL NETWORK, VIEWED FROM n NODE PAIRS [Zₙₙ]

Fig. 10. System interconnection

values of the response currents at the generator terminals give the pu values of the elements of $|Y_{mn}|$, where $|Y_{mn}|$ is the upper right-hand section of Fig. 9(D), i.e.,

$$\|I_m'\| = \|Y_{mn}\| \times \|V_n\|$$

but

$$\|V_n\| = \|Z_{nn}\| \times \|I_n\|$$

where $|Z_{nn}|$ is obtained in the same manner as $|Z_{mm}|$ of the previous section, but disregarding the connection of loads. Therefore $|I_m'| = |Y_{mn}| \times Z_{nn}| \times |I_n|$.

This factor $|Y_{mn}| \times |Z_{nn}|$, can be obtained directly, as shown in Fig. 9(E), by short-circuiting the m generator nodes to neutral and impressing 1-pu current to the load nodes in turn (the other $(n-1)$ load nodes being open-circuited), and measuring $|I_m'|$ at the generator bus bars.

Additional Losses

Additional losses in a system may be due to the connection of more loads, or to connection to another existing system. Fig. 10 shows a system with a number of accessible terminals, to which additional loads (or another system) are to be connected. It can be shown, assuming negligible change at the generator bus bars due to regulation, that the losses in the original system will increase (or decrease) in accordance with the following relationship. The change in losses in the original system is

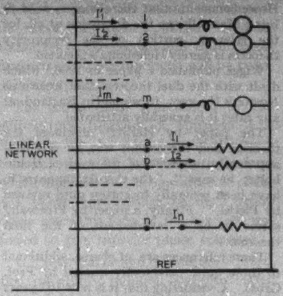


Fig. 11. Effect of load currents on the equivalent circuit of Fig. 9(A).

$$\Delta P = (\|Z_{nn}\| \times \|I_n\|)^* \times \|I_n\|$$

where the asterisk denotes a conjugated quantity, and where $|Z_{nn}|$ is the internal impedance of the original network viewed from the n accessible terminals. It should be noted that the currents through the accessible terminals $|I_n|$ may be positive or negative. The relationship is clearly true since the initial values of $|I_n|$ were zero.

The main point of interest is the case with which the generalization of Thévenin's theorem is applied to give $|I_n|$, since the currents through the terminals 1, 2, .. n are equal to $(|Z_{nn}'| + |Z_{nn}|)^{-1} \times |V_n|_{\mathrm{OC}}$.

This assumption is valid if the generators are not equipped with regulators, and provided the series impedances of the generators are regarded as integral parts of the transmission system, thereby contributing to the losses. However, from the point of view of voltage regulation at the generator bus bars, the effect of connecting new loads is of importance, as follows:

Fig. 11 represents the network where n load accessible terminals are the terminals to which the new loads (or new system) are to be connected; the m generator accessible terminals are those of interest, i.e., those at which the regulation is to be determined. The other generators in the circuit are not of prime interest and are included within the box. Now, as shown above, the current distribution factor $|Y_{mn}| \times |Z_{nn}|$ is measured for the box, thus, $|I_m'| = |Y_{mn}| \times |Z_{nn}| \times |I_n|$. With $|I_m'|$ known, and with the internal impedance matrix of the generators known, the bus-bar voltages may be evaluated.

References

1. Sur un nouveau théorème d'électricité dynamique, L. Thévenin. *Comptes Rendus*, Paris, France, vol. 97, 1883, p. 159.

2. A Useful Network Theorem, J. Millman. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 28, 1940, p. 413.

3. A New Theorem for Active Networks, A. T. Starr. *Journal*, Institution of Electrical Engineers, London, England, vol. 73, 1933.

4. The Generalization of Thévenin's Theorem, C. Adamson, B. D. Nellist. *AIEE CP 59-1189* (available on request).

5. On An Extension of Thévenin's Theorem, H. Ataka. *Philosophical Magazine*, London, England, vol. 25, 1938, pp. 663–66.

6. A Network Theorem, E. E. Zepler. *Wireless Engineer*, London, England, vol. 29, 1952, pp. 44–45.

7. Note On A Useful Extension of Thévenin's Theorem, L. Tasny-Tschiassny. *Proceedings*, Institution of Electrical Engineers, vol. 97, pt. I, 1950, p. 234.

8. Equivalent Circuits—I, Frank M. Starr. *AIEE Transactions*, vol. 51, June 1932, pp. 287–98.

9. Tensorial Analysis of Integrated Transmission Systems—Part I, Gabriel Kron. *Ibid.*, pt. II, vol. 70, 1951, pp. 1239–48; Part II, *Ibid.*, pt. III (*Power Apparatus and Systems*), vol. 71, Jan. 1952, pp. 505–12; Part III, *Ibid.*, Oct., pp. 814–22; Part IV, *Ibid.*, vol. 72, Aug. 1953, pp. 827–39.

10. Tensor Analysis of Networks, Gabriel Kron. John Wiley & Sons, Inc., New York, N.Y., 1939.

◆

Discussion

Eric T. B. Gross (Illinois Institute of Technology, Chicago, Ill.): Since the authors deal with a number of useful network theorems, it may be of interest to add a few historical comments: The "Thévenin" theorem was made known 4 years before L. Thévenin (1857–1926) was born; the "principle of the equivalent generator" was first formulated in 1853 by Helmholtz (1821–1894), but never given his name because it must have been considered a minor contribution of this famous physicist. Even if Thévenin did not know of Helmholtz's earlier work when he wrote his article 30 years later, in 1883, those in this country who "discovered" Thévenin some 40 to 50 years ago would have found from a study of the history of circuit theorems, as contained in many European textbooks, who really originated what they then named "Thévenin's" theorem.

Thévenin published his total findings, a dry statement of five paragraphs, without any proofs, in 1883; this paper gives nothing but the procedure of establishing the equivalent generator (see reference 1 of the paper). Helmholtz' paper, however, is a comprehensive analytical investigation which includes the rigorous proof.[1]

The "useful network theorem" given in the authors' reference 2) is also contained in an earlier work, published in 1921 by Willheim[2] in very general form and including the consideration of an electrical center of gravity. This theorem was later used by many authors, and restated by Willheim in two subsequent publications.[3,4]

References

1. Ueber einige Gesetze der Vertheilung elektrischer Stroeme in koerperlichen Leitern mit Anwendung auf die thierisch-elektrischen Versuche, H. Helmholtz. *Annalen der Physik und Chemie*, Leipzig, Germany, vol. 89, 1853, p. 213.

2. Die allgemeinen Loeschbedingungen fur Erdschluss-Schutzeinrichtungen, R. Willheim. *Elektrotechnik und Maschinenbau*, Vienna, Austria, vol. 39, 1921, p. 137.

3. Das Erdschlussproblem in Hochspannungsnetzen (book), R. Willheim. Julius Springer, Berlin, Germany, 1936, pp. 20–24, 222–23.

4. Neutral Grounding in High-Voltage Transmission (book), R. Willheim, M. Waters. Elsevier Publishing Company, London, England, and New York, N. Y., 1956, pp. 23–29, 412–15.

5. Of What Value is History? I. E. Mouromtseff. *Electrical Engineering*, vol. 68, no. 11, Nov. 1949, pp. 945–48.

Colin Adamson and **S. M. El-Sobki**: The authors are grateful to Prof. Gross for emphasizing that the theorem attributed to Thévenin, was, in fact, formulated by Helmholtz in 1853. It may be of interest to note that the question of the credit due to Helmholtz was dealt with, with finality by Howe in 1943,[1] in an article of considerable interest in that it gives in detail the long succession of authors who at various times after its original formulation, "discovered" the Helmholtz (Thévenin) theorem; see, for example, Pomey.[2] Pleijel, of the Swedish Telegraphic Department, had read a paper at the International Conference of Post and Telegraphs in Paris in September 1910, and

Pomey states that Pleijel has used Thévenin's theorem together with a lemma of his own. Howe points out that the lemma is unnecessary as it follows directly from Helmholtz's theorem.

Chapman applied Thévenin's theorem to alternating current networks in 1923.[3] No references were given for earlier papers on the subject. In 1927, Wenner,[4] published a paper, reprinted with documents from a Scientific paper of the United States Bureau of Standards, which restated Helmholtz's theorem and supplied a great many references, including Helmholz but not Thévenin.

The theorem of Helmholtz (Thévenin) was further described in 1928 by V. Jenkin in the pages of *Revue Generale de l'Électricité*. A further statement of the same theorem was made in 1933 by J. G. Brainard in the *Proceedings* of the Institute of Radio Engineers: Brainard called the theorem Thévenin's but added, "Sometimes called Pollard's theorem."

A more substantial contribution was published in 1935 by Van den Meersche.[5] Howe comments that there was no need to bring Maxwell into the discussion at all, for the relevant portion of the reciprocity theorem is merely Helmholtz's theorem.

Wigge published a paper in 1936,[6] which dealt with the duel theorem that known as Norton's theorem, to whom (in England, at any rate) it is generally attributed.

The paper by Herlitz[7] discusses the theorem and the author was, in this case apparently, unaware of the work of Helmholtz; he says "...the theorem appears to have been generally forgotten until revived by Pleijel." Finally, a paper by Freeman,[8] seems to be a restatement of the dual theorem.

These references are, of course, additional to those quite pertinently quoted by Prof. Gross. Considering that it is now 107 years since Helmholtz first formulated the theorem which does not usually bear his name, it is high time that this situation was corrected. It is unfortunate, indeed, that the theorem is nearly always known as Thévenin's theorem by practicing engineers.

REFERENCES

1. THE MAKE AND BREAK NETWORK THEOREM OF HELMHOLTZ, G. W. O. Howe. *Wireless Engineer*, London, England, vol. 20, no. 238, July 1943, pp. 319–22.

2. THE THEOREM OF PLEIJEL, J. B. Pomey. *Revue Generale de l'Électricité*, Paris, France, vol. 5, Apr. 1919, pp. 622–23.

3. THE CALCULATION OF D. C. AND A. C. NETWORKS, F. T. Chapman. *Electrical Review*, London, England, Mar. 30, 1923.

4. A PRINCIPLE GOVERNING THE DISTRIBUTION OF CURRENT IN SYSTEMS OF LINEAR CONDUCTORS, F. Wenner. *Proceedings*, Physical Society, London, England, 1927.

5. A THEOREM DEDUCED FROM THE GENERALIZED RECIPROCITY THEOREM OF MAXWELL, A. J. van den Meersche. *Revue Génerale de l'Électricité*, Paris, France, vol. 38, Aug. 1935, pp. 259–72.

6. SOME CONSEQUENCES OF THE "ABLEITUNGSSATZ" of Helmholtz (Thévenin's Theorem), H. Wigge. *Archiv fur Elektrotechnik*, Berlin, Germany, 1936.

7. THÉVENIN-PLEIJEL THEOREM, I. Herlitz. *ASEA Journal*, London, England, Apr. 1927.

8. A GENERAL SUPERPOSITION THEOREM OF THE "THÉVENIN" TYPE, G. E. Freeman. *Philosophical Magazine*, London, England, series 7, vol. 33, Jan.-Dec. 1942, pp. 679–85.

Ripple-Type Time-Delay Networks Using Elliptic Functions

J. R. KISEDA
NONMEMBER AIEE

D. J. FORD
ASSOCIATE MEMBER AIEE

ONE METHOD for simulating the approximate effect of transport delay or time delay is through the use of active networks containing electronic operational amplifiers. It is necessary that such networks be used in the field of analog computers so that adequate simulation of problems involving time-delay phenomena can be solved. The thermal and hydraulic problems involved in the design of nuclear reactor power plants form an important class of present-day problems which require such networks.

The purpose of this paper is to present the methods whereby time-delay networks may be designed on the basis of a criterion set up in the frequency domain. The network is selected to have a gain magnitude characteristic which is constant with frequency and a group-delay characteristic (the first derivative of the phase angle with respect to angular frequency) which ripples about the ideal constant value in a prescribed manner. The design methods permit the group delay to assume virtually any ripple-type function. As an example, an elliptic-type function will be studied. Since the ripples are to be considered small in relative magnitude, the phase angle may be assumed to possess a ripple characteristic similar to the group delay.

Historical Background

A number of recent investigators have developed analog computer networks to approximate an ideal time delay or the ideal transport lag. Two studies with which the present work is to be compared are those of Stubbs and Single[1] and Fox.[2] Stubbs and Single produced networks which could be developed from cascaded second-order networks, each of which would have a ripple-type phase characteristic. Their approach was a semigraphical one whereby the variation in the ripple magnitude was considered as certain parameters were varied. The networks of Stubbs and Single produced a phase characteristic which rippled in a somewhat undefined way; the resulting configurations did produce large phase shifts for a given maximum specified phase error. Fox developed circuits with a linear type of "maximally-flat" phase characteristic. These circuits maintained extremely small phase errors in the low-frequency range, with a smooth non-ripple-type phase characteristics throughout the useful frequency range.

Basic Theory

The transfer function of the ideal time-delay device is taken to be $\exp(-sT_d)$ in which s is the complex Laplace transform variable, and T_d is the magnitude of the time delay, or transport delay. For steady-state sinusoidal conditions, this becomes $\exp(-j\omega T_d)$. Appendix I gives a brief proof that the quotient of two infinite-order polynomials represents exactly this ideal steady-state transfer function, provided that the restrictions on the n coefficients implied by equations 21, 23, and 25 are satisfied. If the two series of 21 are terminated at a finite number of terms, then the resulting quotient will approximate the ideal transfer function in a manner dependent upon the restrictions placed on the finite number of A_n, B_n coefficients. Appendix II shows that if the numerator and denominator polynomials are terminated at $n=N$, and if the steady-state gain magnitude is specified as unity for the infinite-frequency range, then the numerator and denominator coefficients must be related as indicated in equation 1.

$$A_{2n}=B_{2n},\ n=0,1,2,\ldots \qquad (1)$$

$$A_{2n-1}=-B_{2n-1},\ n=1,2,3,\ldots$$

When these restrictions are applied to

Paper 59-1193, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted March 23, 1959; made available for printing August 11, 1959.

J. R. KISEDA, formerly with the University of Pittsburgh, Pittsburgh, Pa., is now with International Business Machines Corporation, Yorktown Heights, N. Y., and D. J. FORD is with the University of Pittsburgh.

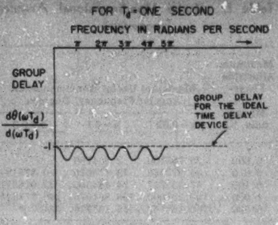


Fig. 1. Ripple-type group delay, curve an elliptic-type function

equation 21, it may be written in the form:

$$e^{-j\omega T_d} \approx \frac{\sum_{n=0}^{N} \cos\left(\frac{n\pi}{2}\right) A_n(\omega T_d)^n - j\sum_{n=0}^{N} \sin\left(\frac{n\pi}{2}\right) A_n(\omega T_d)^n}{\sum_{n=0}^{N} \cos\left(\frac{n\pi}{2}\right) A_n(\omega T_d)^n + j\sum_{n=0}^{N} \sin\left(\frac{n\pi}{2}\right) A_n(\omega T_d)^n} \qquad (2)$$

A magnitude response of unity for all frequencies is typical of all-pass structures, hence equation 2 may be thought of as representing that class of networks.

The object of this paper is to present a general method by which the A_n coefficients of equation 2 can be selected so as to produce a prescribed ripple-type group delay characteristic. This is accomplished by setting an exact (or approximate) form of the desired ripple function equal to the group delay calculated from the right-hand side of equation 2.

In doing this, it is assumed that the magnitude of the ripple is relatively small, so that there will be a somewhat similar ripple in the phase characteristic. The exact shape of the ripple in the phase characteristic is not treated as a design quantity directly, since it was found more convenient to manipulate the group-delay characteristic.

Three main types of ripple functions could be classified as: 1. an equal ripple and periodic function; 2. an equal ripple and nonperiodic function; and 3. a nonequal ripple and nonperiodic function. The function chosen to exhibit the first classification was an elliptic-type function. If $\theta(\omega T_d)$ is the phase angle of the right-hand side of equation 2, the group delay can be set equal to an elliptic-type function in the following manner:

$$\frac{d\theta(\omega T_d)}{d(\omega T_d)} = \frac{-1}{\sqrt{1-k^2 \sin^2 \omega T_d}} \qquad (3)$$

where k controls the maximum value of the curve's deviation from the -1 value. Examples of the second and third classifications of ripple functions will be given in a later paper.

The group delay represented by the elliptic-type function of equation 3 is shown in Fig. 1. This form of the function was selected in preference to the form which oscillates on the opposite side of the -1 value representing the ideal group delay since this form of the function was found to give superior phase response.

In the actual design techniques utilized, the true elliptic function is approximated by a finite number of terms of its binomial expansion. The resulting phase characteristic ripples about the ideal linear-phase characteristic in a manner which produces both positive and negative phase errors. (The phase error is defined as the difference between the phase response of the ideal time-delay transfer function and the calculated theoretical values of the approximation.)

Second-Order Ripple-Phase Approximation

Let N take on the value 2 in equation 2. The resulting transfer function will be a a second-order approximation to the ideal time-delay transfer function, exp$(-j\omega T_d)$. If the A_0 coefficient is factored out of the numerator and the denominator of equation 2, the remaining coefficients could be designed A_1' and A_2'. If A_1' is set equal to KA_2' and the primes are dropped for convenience, the resulting second-order approximation function can be written as:

$$e^{j\omega T} \approx \frac{1-jKA_2(\omega T_a)-A_2(\omega T_a)^2}{1+jKA_2(\omega T_d)-A_2(\omega T_a)^2} \qquad (4)$$

The magnitude of the right-hand side of equation 4 is unity for all real frequencies in a manner typical of all-pass network structures. The phase angle of the right-hand side of equation 4 can be shown to be:

$$\theta(\omega T_d) = -2\tan^{-1}\left[\frac{KA_2(\omega T_d)}{1-A_2(\omega T_d)^2}\right] \qquad (5)$$

while the corresponding group delay becomes:

$$\frac{d\theta(\omega T_d)}{d(\omega T_d)} = -2\times \left[\frac{KA_2+KA_2^2(\omega T_d)}{1+(K^2A_2^2-2A_2)(\omega T_d)^2+A_2^2(\omega T_d)^4}\right] \qquad (6)$$

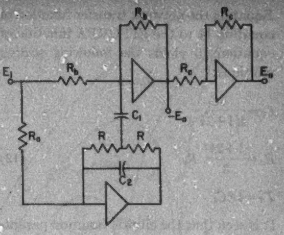


Fig. 2. Second-order ripple-phase circuit

In order to require that the group delay assume an elliptic function form, equation 6 is set equal to equation 3. Comparing these equations, it does not seem possible to require the group delay of the second-order system to fit the elliptic function exactly. Expanding the denominator of the elliptic function in a binomial series and retaining only the first three terms (which requires that k be preferably much less than unity for a reasonably accurate approximation), equations 3 and 6 can be combined to show that:

$$-1 = \left[\frac{-2KA_2+(k^2KA_2-2KA_2^2)(\omega T_d)^2+\left(k^2KA_2^2-\frac{k^2KA_2}{3}\right)(\omega T_d)^4}{1+(KA_2^2-2A_2)(\omega T_d)^2+A_2^2(\omega T_d)^4}\right] \qquad (7)$$

from which the parameters A_2 and K can be determined in terms of the elliptic function parameter k as:

$$A_2 = \frac{1+2k^2}{12} \qquad (8)$$

$$K = \frac{6}{1+2k^2} \qquad (9)$$

This provides the means whereby the elliptic function second-order ripple-phase approximation (EFSORPA) transfer function can be written as:

$$e^{-j\omega T_d} \approx \frac{1-j\frac{1}{2}(\omega T_d)-\left(\frac{1+2k^2}{12}\right)(\omega T_d)^2}{1+j\frac{1}{2}(\omega T_d)-\left(\frac{1+2k^2}{12}\right)(\omega T_d)^2} \qquad (10)$$

Equation 10 completes the mathematical development of the transfer function for the second-order approximation.

One method for physically realizing the transfer function of equation 10 is through use of the analog computer circuit given in Fig. 2. The transfer function of the network in terms of its parameters can be shown to be:

$$\frac{\vec{E}_o}{\vec{E}_i} = \frac{1+j\omega\left(2RC_2-\frac{RR_bC_1}{R_a}\right)-\omega^2R^2C_1C_2}{1+j2\omega RC_2-\omega^2R^2C_1C_2} \qquad (11)$$

Equating the network transfer function of equation 11 to the EFSORPA function of equation 10 yields the following scaling relations:

$$C_2=\frac{3C_1}{4(1+2k^2)}$$

$$R_a=\frac{1+2k^2}{3}\,R_b \qquad (12)$$

$$T_d=4RC_2$$

It is seen that the elliptic function parameter k fixes the ratio of C_1 to C_2 and R_a to R_b. The selection of C_2 allows the time delay T_d to be a linear function of the resistance R.

A set of curves giving the magnitude of the phase error as a function of the nondimensional angular frequency, ωT_d, for various values of k is drawn to semilog coordinates as Fig. 3. The curves of Fig. 3 possess a characteristic which could be termed a bandwidth. This bandwidth could be defined as the frequency range from zero to the highest value for which the per-cent phase-error magnitude equals the value attained at the "hump" or first peak in the curve. By changing the elliptic function parameter k, the bandwidth can be increased or decreased, though the maximum error within the bandwidth varies correspondingly.

For each value of the per-cent phase-error magnitude, there exists a maximum frequency range over which the error is equal to or less than this value. This frequency range varies with the elliptic function k, for a given per-cent phase-error magnitude. Table I gives some calculated values in the relationship between the per-cent phase-error magnitude and the maximum useful frequency range for this error for three different values of k.

It can be shown that the maximum possible phase angle of the EFSORPA function (and the equivalent analog computer network of Fig. 2) ripples slowly about the ideal curve until the maximum usable frequency is reached. Beyond this point the phase angle departs appreciably from the ideal curve and approaches its maximum value of 2π asymptotically.

Experimental Results for EFSORPA Network

To test the feasibility and accuracy of the EFSORPA function development, a set of sinusoidal magnitude and phase-shift tests were conducted on the network of Fig. 2. The measuring equipment available at the time of these tests required that the frequency range be restricted to 20 cps (cycles per second) and above. Some preliminary runs indicated that the useful frequency range for 1% error was about 140 degrees. With this information, a phase shift of 90 degrees was set at 60 cps, and the time delay T_d fixed accordingly. The other parameter values were determined by the scaling relationships of equation 12 and by the selection of arbitrary reasonable values for the small-size analog computer equipment available. The calculated values for the tests are:

$T_d=4.16667\times10^{-3}$ seconds
$C_1=1.0\times10^{-7}$ farads
$R_b=1.0\times10^{6}$ ohms

k	C_2, Farads	R, Ohms	R_a, Ohms
0.02	7.494×10^{-8}	13,900	333,600
0.10	7.353×10^{-8}	14,167	340,000
0.20	6.944×10^{-8}	15,000	360,000

The input and output voltages for the operational amplifier network were measured in the usual way by vacuum-tube

Table I. Useful Nondimensional Angular Frequency*

Maximum Allowable Phase-Error Magnitude, %	Maximum Useful Nondimensional Angular Frequency, Degrees		
	k=0.02	k=0.1	k=0.2
0.002	25.783100	2.864789	
0.010	34.377466	13.178029	6.875493
0.100	57.295777	74.484509	22.918311
1.000	100.267609	108.861976	140.374645
5.000	160.428162	166.157736	184.0

* Second-order ripple-phase approximation using elliptic-type functions.

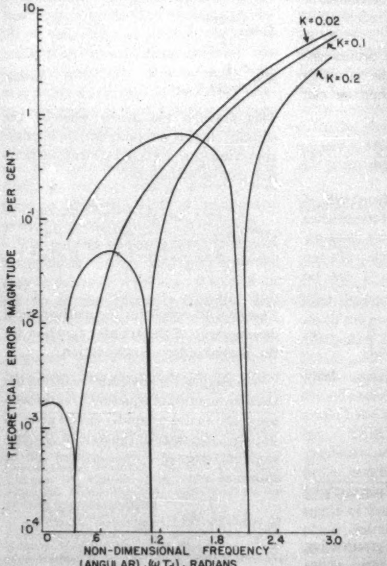


Fig. 3 (left). Theoretical phase-error magnitude in per cent versus nondimensional angular frequency for EFSORPA. All errors plotted to left of tip of dip are minus values

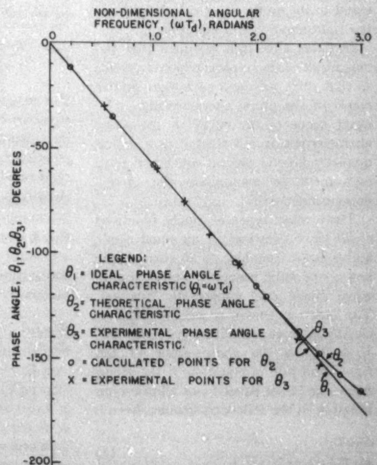


Fig. 4 (right). Experimental phase angle (or phase shift) and theoretical phase angle (or phase shift) versus nondimensional angular frequency for EFSORPA

Vol.
Pt. 1
1959

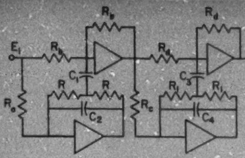


Fig. 5 (left). Fourth-order ripple-phase circuit

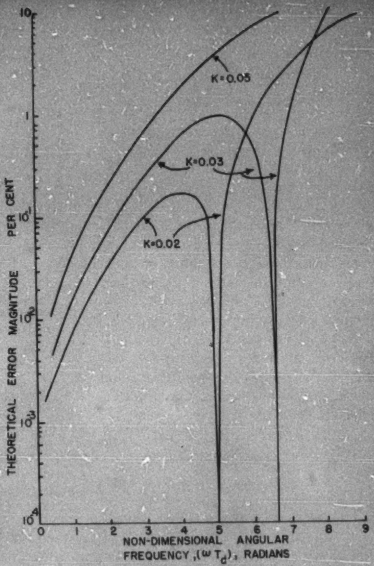


Fig. 6. Theoretical phase-error magnitude in per cent versus nondimensional angular frequency for EFFORPA. All errors plotted to left of tip of dip are minus values

voltmeters which were calibrated to an accuracy of approximately ±1%. The gain magnitude was unity throughout the frequency range tested. The phase shift between input and output was measured by the 3-voltmeter method.

The experimental phase-shift data is given in the curves of Fig. 4 by the series of *x* marks. As can be seen, the measured values agree very well with the calculated curve. The maximum deviation was 1.3%, which occurred at 60 cps, and was probably due to hum voltage pickup within the circuit. Only the data for *k* equal to 0.2 is given, since the data for the other values of *k* agreed with the calculated values within the same accuracy limits.

Fourth-Order Ripple-Phase Approximation

To develop a fourth-order ripple-phase approximation, the *N* of equation 2 is set equal to four. The general procedure beyond this point is basically the same as in the second-order approximation. However, one or two slight variations on the fundamental method are possible. It might seem desirable to cascade two identical second-order linear-phase approximation networks to form a fourth-order network. This would double the phase angle at each frequency, but it would also double the error, so that the percentage error at each frequency would remain the same as in the second-order case. If one considers a fixed time delay, T_d, this is equivalent to saying that a fourth-order approximation formed in this way would produce precisely twice the useful frequency range, for a given magnitude phase error that the second-order approximation would produce. If, on the other hand, one considers cascading two different second-order ripple-phase approximation functions, it is possible to obtain more than twice the useful frequency range for a given magnitude phase error. This method will be illustrated in the following.

Let the basic polynomial approximation function of equation 2, with *N* equal to four, be broken into two different second-order factors as:

$$\epsilon^{-j\omega T_d} \approx \frac{A_0 - jA_1(\omega T_d) - A_2(\omega T_d)^2}{A_0 + jA_1(\omega T_d) - A_2(\omega T_d)^2} \times \frac{B_0 - jB_1(\omega T_d) - B_2(\omega T_d)^2}{B_0 + jB_1(\omega T_d) - B_2(\omega T_d)^2} \quad (13)$$

Dividing by the leading coefficient and defining new coefficients and constants in the same manner as in the second-order approximation gives:

$$\epsilon^{-j\omega T_d} \approx \frac{1 - jKA_2(\omega T_d) - A_2(\omega T_d)^2}{1 + jKA_2(\omega T_d) - A_2(\omega T_d)^2} \times \frac{1 - jLMA_2(\omega T_d) - MA_2(\omega T_d)^2}{1 + jLMA_2(\omega T_d) - MA_2(\omega T_d)^2} \quad (14)$$

From this, the group delay can be shown to be:

$$\frac{d\theta(\omega T_d)}{d(\omega T_d)} = -2A_2 \times \left[\frac{\begin{array}{l}(K+LM)+\\ (\omega T_d)^2(A_2)(K^2LMA_2+KL^2M^2A_2-\\ \quad 2KM+K+LM^2-2LM)+\\ (\omega T_d)^4(A_2^2M)(K^2LMA_2+KL^2MA_2-\\ \quad 2K+KM+L-2LM)+\\ \quad (\omega T_d)^6(A_2^3M^2)(K+L)\end{array}}{\begin{array}{l}1+(\omega T_d)^2(A_2)(K^2A_2-\\ \quad 2+L^2M^2A_2-2M)+\\ (\omega T_d)^4(A_2^2)(K^2L^2M^2A_2^2-2L^2M^2A_2-\\ \quad 2K^2MA_2+M^2+4M+1)+\\ \quad (\omega T_d)^6(A_2^3M)(K^2MA_2+L^2MA_2-\\ \quad 2M-2)+(\omega T_d)^8(M^2A_2^4)\end{array}}\right] \quad (15)$$

This group delay is set equal to the desired elliptic function (see equation 3), the denominator of which is expanded in a binomial series as before. Retaining the first five terms of this expansion gives:

$$\frac{d\theta(\omega T_d)}{d(\omega T_d)} = \frac{1}{1 - \frac{k^2}{2}\left[(\omega T_d)^2 - \frac{(\omega T_d)^4}{3} + \frac{2}{45}(\omega T_d)^6 - \frac{(\omega T_d)^8}{315}\right]} \quad (16)$$

Thus equating equation 16 to equation 15 gives implicitly the required relationship between the coefficient parameters of the general time-delay approximation function, A_2, *K*, *L*, *M*, and the *k* parameter of the elliptic function. This set of equations can be written as:

$$0 = KA + LMA - 1/2$$

$$0 = K^2A^2 - 3KA + L^2M^2A^2 - 3LM^2A + \frac{k^2}{2}(K+LM)$$

$$\begin{aligned}0 = {} & K^2L^2M^2A^4 + K^2L^2M^3A^4 - 2K^2MA^3 - 3K^2LM^2A^3 - \\ & 3KL^2M^2A^3 + 6KMA^2 + KA^2 - 2L^2M^2A^3 + LM^3A^2 + 6LM^2A^2 - \\ & \frac{k^2}{6}(K+LM) + \frac{k^2}{2}[(K^2LMA^2 + KL^2M^2A^2 - 2KMA + \\ & KA + LM^2A - 2LMA)]\end{aligned}$$

$$\begin{aligned}0=\;&K^2M^2A^4+K^2LM^2A^4+\\&KL^2M^2A^4-3KM^2A^3-2KMA^3+\\&L^2M^2A^4-2LM^2A^3-3LM^2A^3+\\&3KM^2A^4+3LM^2A^4-\\&\frac{k^2}{10}[K^2LMA+KL^2MA^2-\\&2KMA+KA+LM^2A-2LMA]+\\&\frac{11k^2}{630}\,[K+LM]-\frac{3k^2}{2}\times\\&[KM^2A^3+LM^2A^3]\end{aligned}\qquad(17)$$

where the 2 subscript on the A has been dropped for convenience in notation.

Solution of the four simultaneous nonlinear algebraic equations 17 was performed by use of the Newton-Raphson method with Fortran programming techniques applied to an IBM (International Business Machines Corporation) *704* digital computer. Solutions for k equal to 0.02, 0.03, 0.04, and 0.05 were determined. These are given as:

	Elliptic Parameter k			
	0.02	0.03	0.04	0.05
K	10.888727	10.382075	9.917709	9.480453
L	7.812266	7.131397	6.304980	5.368634
M	0.712903	0.668802	0.644128	0.637577
A	0.030380	0.033000	0.035768	0.038750

The substitution of any set of these values into equation 14 yields the final form of the elliptic function fourth-order ripple-phase approximation, or EFFORPA, function for a given k. For $k=0.02$, this can be written as:

$$\epsilon^{-j\omega T_d}\approx\left[\frac{1-j0.330800(\omega T_d)-0.030380(\omega T_d)^2}{1+j0.330800(\omega T_d)-0.030380(\omega T_d)^2}\right]\times\left[\frac{1-j0.169198(\omega T_d)-0.021658(\omega T_d)^2}{1+j0.169198(\omega T_d)-0.021658(\omega T_d)^2}\right]\qquad(18)$$

When k was increased above 0.05, it was found to be impossible to solve the the set of equations 17 with K, L, M, and A restricted to real values.

An analog computer circuit which will realize the transfer function given in equation 14 is shown in Fig. 5. The transfer function of this circuit is:

$$\frac{\vec{E}_0}{\vec{E}_i}=\left[\frac{1+j\omega\bigg(2RC_2-\dfrac{RR_bC_1}{R_a}-\omega^2R^2C_1C_2}{1+j^2\omega RC_2-\omega^2R^2C_1C_2}\right]\times\left[\frac{1+j\omega\left(2R_1C_4-\dfrac{R_1R_dC_3}{R_c}\right)-\omega^2R_1{}^2C_3C_4}{1+j2\omega R_1C_4-\omega^2R_1{}^2C_3C_4}\right]\qquad(19)$$

Since a general solution to the set of equations 17 could not be found, a general set of scaling relationships in terms of k could not be developed either. In a specific case, these can be evaluated. For $k=0.02$, the following can be shown to be true:

$$\begin{aligned}T_d&=6.045949RC_2\\C_2&=0.277623R_b\\C_4&=0.33045C_3\\R_c&=0.756532R_d\end{aligned}\qquad(20)$$

The calculated phase-shift error magnitude in per cent is given in Fig. 6 as a function of nondimensional angular frequency ωT_d. Table II gives the maximum usable nondimensional angular frequency for several values of the phase error magnitude for the data shown in Fig. 6. Comparison of this table with Table I shows that the maximum useful frequency range of the fourth-order network is appreciably more than twice that attainable for the second-order network for the same k. For example, if the phase error is restricted to 1% or less, the maximum usable frequency range of the fourth-order network is 3.25 times that attainable with the second-order configuration for the same k value of 0.02.

The fact that the physical realization of the fourth-order configuration requires only four operational amplifiers as compared to the three amplifiers which were required in the second-order network seems to make the fourth-order network even more attractive. Other physical realizations of the defining equations may be possible which would render this comparison apparently less favorable. Stubbs and Single were able to produce a fourth-order network using only three operational amplifiers and a number of additional components. Questions of convenience in operation and adjustment might make the 4-amplifier configuration presented herein superior to the 3-amplifier configuration of Stubbs and Single.

A comparison of the maximum useful frequency ranges for specified phase-error magnitudes for the Stubbs and Single network, the Fox maximally flat phase-response network, and the elliptic-type ripple-phase function networks is made in Table III. For the extremely low values of phase-error magnitude, the Fox circuit is superior to all other types while for the larger values of phase-error magnitude, the Stubbs and Single circuit gives from 5 to 10% larger useful frequency range than the particular EFFORPA cases calculated. Since the EFFORPA network response changes markedly with the parameter k, it is felt that larger useful frequency ranges could be obtained through further calculations. This is perhaps more evident from Table II and Fig. 6. For example, it is felt that an optimum value for the useful frequency range at a phase-error magnitude of 5% could be found for a k value in the vicinity of 0.035. It is further believed that this optimum value would compare favorably with the range attained by the Stubbs and Single network. By varying k, the optimum useful frequency range could be obtained for each value of phase-error magnitude of interest. A chart or set of curves could be drawn to portray this information easily. Experimentally, a similar set of data could be obtained by varying the resistance and capacitance values of Fig. 5 to obtain a variation in k. The lack of a general solution for k in terms of the network parameters would hamper this method, but the ranges of

Table II. Maximum Useful Nondimensional Angular Frequency*

Maximum Allowable Phase-Error Magnitude, %	Maximum Useful Nondimensional Angular Frequency, Degrees			
	k=0.02	k=0.03	k=0.04	k=0.05
0.001	2.864789	2.864789	3.437747	
0.010	65.890143	45.836622	34.377466	25.783100
0.100	177.616888	128.915493	103.132398	85.943665
1.000	352.368881	401.070248	223.453495	194.805616
5.000	444.042042	475.554687	509.932121	309.397087

* Fourth-order ripple-phase approximation using elliptic functions.

Table III. Maximum Useful Nondimensional Angular Frequencies*

Maximum Allowable Phase-Error Magnitude, %	Maximum Useful Nondimensional Angular Frequency, Degrees				
	Fox Circuit	Stubbs and Single Circuit	EFFORPA		
			k=0.02	k=0.03	k=0.04
0.0015	120.32114				
0.0100	160.42818	28.6	65.89014	45.83662	34.37747
0.1000	216.57805	45.8	177.61689	128.91549	103.13240
1.0000	304.81355	453.0	352.36888	401.07025	223.45350
5.0000	401.07046	532.0	444.04204	475.55469	509.93212

* For various fourth-order approximation networks.

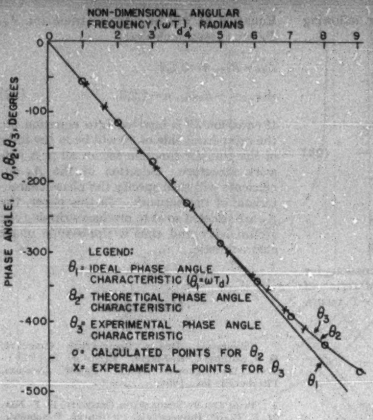


Fig. 7. Experimental phase angle (or phase shift) and theoretical phase angle (or phase shift) versus nondimensional angular frequency for EFFORPA

values to be investigated is made reasonably narrow by the solutions presented in the foregoing.

From a partical viewpoint, only those frequency ranges, in the vicinity of the 0.10, 1, and 5% phase-error magnitude values of Table III would probably be significant. For the EFFORPA network, the variation in the frequency range with k for these phase-error values again points to the wide range of characteristics obtainable.

Experimental Results for EFFORPA Network

The EFFORPA network with k equal to 0.02 was tested experimentally by the same methods as those used to test the EFSORPA network. The time delay was selected by setting the 60 cps phase shift at 180 degrees. The scaling relations were developed in a manner similar to that indicated for the EFSORPA case. The numerical values used are given as:

$T_d=8.33333\times10^{-3}$ seconds
$C_1=1.0\times10^{-7}$ farads
$C_2=9.005\times10^{-8}$ farads
$C_3=1.0\times10^{-7}$ farads
$C_4=3.3045\times10^{-7}$ farads
$R_a=277{,}623$ ohms
$R_b=1{,}000{,}000$ ohms
$R_c=756{,}432$ ohms
$R_d=1{,}000{,}000$ ohms
$R=15{,}306$ ohms
$R_1=21{,}334$ ohms

The experimental phase-shift points are given in Fig. 7. The same very close agreement of the experimental points with the calculated curves was obtained in this case as was found for the EFSORPA network. The magnitude ratio was constant at unity through out the frequency range tested.

Conclusions

A method of designing approximate time-delay simulation network in the frequency domain has been presented. The method utilizes an expansion of the steady-state sinusoidal transfer function of an ideal time-delay device into the quotient of two polynomials in ωT_d. By taking a finite number of terms of the expansion, an approximation to the ideal time-delay device is obtained. If the corresponding coefficients in the numerator and denominator are set equal (except for the signs of some terms), the quotient will possess first, a constant unity-gain for all positive frequencies which is characteristic of all-pass structures, and second, a phase characteristic which may be adjusted at will. By setting the group-delay equal to a desired ripple-type function, it is possible to cause the phase characteristic to ripple about the ideal linear-phase curve. The details of the development were illustrated for an elliptic-type ripple function in second- and fourth-order transfer-function configurations. The method has been applied to other ripple-type functions but the results were not presented here.

No attempt was made to obtain the maximum possible usable frequency range for the elliptic-type ripple-function approximation networks under conditions of various specified phase-error magnitudes. The general range of k for this optimum condition was indicated for the 1 and 5% phase-error magnitude cases in the calculated data. One calculated case indicated that it was possible to obtain considerably more than twice the usable frequency range with a fourth-order approximation network, than was possible with a second-order network for a given phase-error magnitude and a given elliptic-function parameter value.

It is believed that the design methods presented in this paper give a methodical way of developing a large class of ripple-type time-delay simulation networks.

Appendix I. Validity of Approximation for $e^{-j\omega T_d}$

The expansion chosen for the ideal time-delay transfer function in this paper is

$$\epsilon^{-j\omega T_d}=\frac{\sum_{n=0}^{\infty}(-j)^nA_n(\omega T_d)^n}{\sum_{n=0}^{\infty}(-j)^nB_n(\omega T_d)^n} \tag{21}$$

which could be expanded to:

$$\epsilon^{-j\omega T_d}=\frac{A_0-jA_1(\omega T_d)-A_2(\omega T_d)^2+jA_3(\omega T_d)^3+A_4(\omega T_d)^4+\ldots}{B_0-jB_1(\omega T_d)-B_2(\omega T_d)^2+jB_3(\omega T_d)^3+B_4(\omega T_d)^4+\ldots} \tag{22}$$

A judicious choice of the coefficients of $(\omega T_d)^n$ in equation 22 would permit it to be expressed as:

$$\epsilon^{-j\omega T_d}=\lim_{n\to\infty}\left[\frac{1-\frac{ja_1}{n}(\omega T_d)-\frac{a_2}{n^2}(\omega T_d)^2}{1-\frac{jb_1}{n}(\omega T_d)-\frac{b_2}{n^2}(\omega T_d)^2}\right]^n \tag{23}$$

It can be shown that

$$\lim_{n\to\infty}\left[\frac{1-\frac{ja_1}{n}(\omega T_d)-\frac{a_2}{n^2}(\omega T_d)^2}{1-\frac{jb_1}{n}(\omega T_d)-\frac{b_2}{n^2}(\omega T_d)^2}\right]^n=\epsilon^{-j(a_1+b_1)(\omega T_d)} \tag{24}$$

If a_1 and b_1 were restricted by the following relationship:

$$a_1+b_1=1 \tag{25}$$

then:

$$\lim_{n\to\infty}\left[\frac{1-\frac{ja_1}{n}(\omega T_d)-\frac{a_2}{n^2}(\omega T_d)^2}{1-\frac{jb_1}{n}(\omega T_d)-\frac{b_2}{n^2}(\omega T_d)^2}\right]^n=\epsilon^{-j\omega T_d} \tag{26}$$

Thus it has been shown that

$$\epsilon^{-j\omega T_d}=\frac{\sum_{n=0}^{\infty}(-j)^nA_n(\omega T_d)^n}{\sum_{n=0}^{\infty}(-j)^nB_n(\omega T_d)^n} \tag{27}$$

provided that the restrictions imposed on the coefficients of equation 21 by equations 23 and 25 are satisfied.

It is to be noted this illustrates that the correct choice of the A_n and B_n coefficients of the right-hand side of equation 21 allows one to say that this quotient of two infinite-order polynomials becomes exactly the exponential function, exp $(-j\omega T_d)$. Since the angle of exp $(-j\omega T_d)$ is ωT_d, a completely linear function of ω, the correct choice of A_n, B_n in equation 21 will produce precisely this phase characteristic. The objective of this paper is to show that by using a truncated form of the right-hand side of equation 21 and by setting slightly different restrictions upon the (finite number of) A_n, B_n coefficients, a phase characteristic can be produced which will ripple about this exact phase characteristic, exp $(-j\omega T_d)$ in a prescribed way over a finite-frequency range while the magnitude characteristic can be made to be precisely constant at unity for this same finite-frequency range (as well as, incidently, the entire infinite-frequency range).

Appendix II. Restrictions on Coefficients of All-Pass Network Transfer Function Approximation to Ideal Time Delay

It has been shown in Appendix I that:

$$\epsilon^{-j\omega T_d} = \frac{A_0 - jA_1(\omega T_d) - A_2(\omega T_d)^2 + jA_3(\omega T_d)^3 + A_4(\omega T_d)^4 + \ldots}{B_0 - jB_1(\omega T_d) - B_2(\omega T_d)^2 + jB_3(\omega T_d)^3 + B_4(\omega T_d)^4 + \ldots} \qquad (28)$$

was a valid approximation for $\epsilon^{-j\omega T_d}$ provided that the restrictions imposed on the coefficients of $(\omega T_d)^n$ of equation 28 by equatious 23 and 25 were satisfied. Equation 28 can be written in the following equivalent form:

$$\epsilon^{-j\omega T_d} = \frac{\sum_{n=0}^{\infty} \cos\left(\frac{n\pi}{2}\right) A_n(\omega T_d)^n - j\sum_{n=0}^{\infty} \sin\left(\frac{n\pi}{2}\right) A_n(\omega T_d)^n}{\sum_{n=0}^{\infty} \cos\left(\frac{n\pi}{2}\right) B_n(\omega T_d)^n - j\sum_{n=0}^{\infty} \sin\left(\frac{n\pi}{2}\right) B_n(\omega T_d)^n} \qquad (29)$$

To form an approximate expression for exp $(-j\omega T_d)$, a finite number of terms of equation 28 is selected, say $N-1$ terms. Then equation 28 becomes:

$$\epsilon^{-j\omega C_d} \approx \left| \frac{\sum_{n=0}^{N} \cos\left(\frac{n\pi}{2}\right) A_n(\omega T_d)^n - j\sum_{n=0}^{N} \sin\left(\frac{n\pi}{2}\right) A_n(\omega T_d)^n}{\sum_{n=0}^{N} \cos\left(\frac{n\pi}{2}\right) B_n(\omega T_d)^n - j\sum_{n=0}^{N} \sin\left(\frac{n\pi}{2}\right) B_n(\omega T_d)^n} \right. \qquad (30)$$

If the magnitude of equation 30 is set equal to unity for all frequencies, then:

$$\left\{ \frac{\left[\sum_{n=1}^{N} \cos\left(\frac{n\pi}{2}\right) A_n(\omega T_d)^n\right]^2 + \left[\sum_{n=0}^{N} \sin\left(\frac{n\pi}{2}\right) A_n(\omega T_d)^h\right]^2}{\left[\sum_{n=1}^{N} \cos\left(\frac{n\pi}{2}\right) B_n(\omega T_d)^n\right]^2 + \left[\sum_{n=1}^{N} \sin\left(\frac{n\pi}{2}\right) B_n(\omega T_a)^n\right]^2} \right\}^{1/2} = 1 \qquad (31)$$

Equation 31 will be satisfied when the A_n, B_n are chosen such that:

$$A_{2n} = B_{2n},\ n = 0,1,2,\ldots$$

$$A_{2n-1} = -B_{2n-1},\ n = 1,2,3,\ldots \qquad (32)$$

If equation 32 is inserted into equation 30, the right-hand side of 30 will be in the form of the transfer function for an all-pass network structure. Selection of the A_n coefficients will then specify the phase characteristic of the network. In this paper, the A_n are selected so as to produce a ripple-type group delay and thus a ripple-type phase characteristic.

References

1. Transport Delay Simulation Circuits, G. S. Stubbs, C. H. Single. *Report no. WAPD-T-38m*, Westinghouse Atomic Power Division, Pittsburgh, Pa., 1954.

2. Time Delay Simulation Circuits, D. F. Fox. *M. S. Thesis*, University of Pittsburgh, Pittsburgh, Pa., 1957, pp. 24, 26, 86–89, 93–98.

3. Time Delay Networks for an Analog Computer, W. J. Cunningham. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., vol. PGEC-3, Dec. 1954, pp. 16–18.

Subarea Determination of the Capacitance of Cocentric Annular-Plate Capacitors

DANIEL K. REITAN
MEMBER AIEE

THOMAS J. HIGGINS
FELLOW AIEE

Synopsis: The basic theory of calculation of the capacitance of a capacitor comprised of two cocentric annular plates by the method of (annular) subareas is advanced. Confirming the theory, excellent agreement results between the several available experimentally determined (by electrolytic tank) values for specifically dimensioned capacitors and corresponding calculated values. A family of universal curves is constructed. These afford, by inspection, values of capacitance over the range of geometry of interest in practice.

IN A SERIES of recent papers Aikawa, Oshiyama, Sano, Nakaya,[1–3] Sato, and Ataka[4–6] discuss the capacitance of a capacitor consisting of two cocentric circular annular plates, relative to the use of this configuration in the important industrial problem of determining moisture content of various materials by measuring their "effective" dielectric constants. In addition, this problem is of other than purely electrical interest. Knowledge of a parameter analogous to the capacitance is required, for example, in the hydrodynamic or aerodynamic problems associated with the irrotational flow of a fluid through cocentric annular orifices in an infinite barrier,[7] or of the irrotational flow of a fluid around a thin cocentric annular plate,[8] or of a thin cocentric annular plate moving in a viscous fluid in the direction of its normal.[9]

Exact determination of the capacitance of (and such associated aspects as the charge density over) a stated finite plane area is a problem of such analytical difficulty that exact solutions in terms of tabulated functions have been effected to date only for the circular and elliptic discs.[10,11] In consequence it is to be expected that known solutions for a capacitor composed of two finite plane areas would be both more restricted in geometry and more difficult to effect in analysis. Such is actually the case. It would seem that solutions are known only for capacitors composed of two coaxial[12–23] or two coplanar[24] identical circular plates,

---

Paper 59-1186, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted July 8, 1959; made available for printing August 11, 1959.

Daniel K. Reitan and Thomas J. Higgins are with the University of Wisconsin, Madison, Wis.

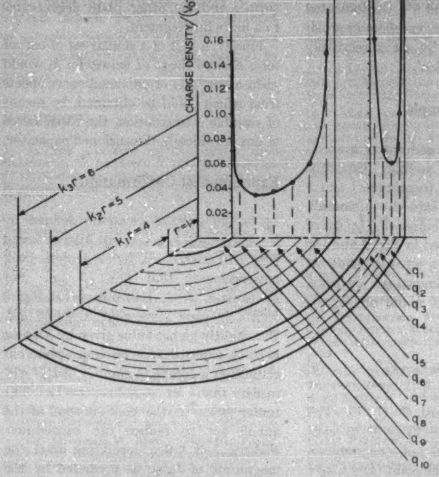

CHARGE DENSITY/(V_0-V_0')
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
$k_3r=6$
$k_2r=5$
$k_1r=4$
$r=1$
q_1
q_2
q_3
q_4
q_5
q_6
q_7
q_8
q_9
q_{10}

Fig. 1. One quadrant of a specific annular capacitor divided into 10 subannuli

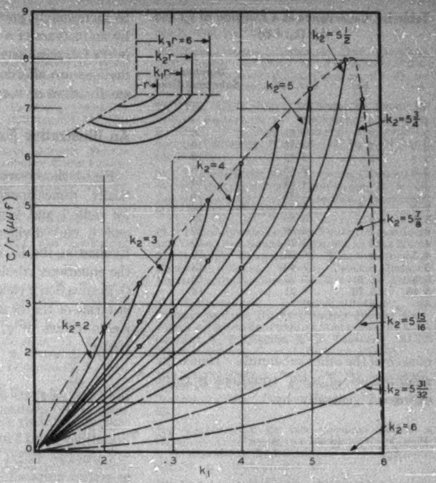


Fig. 3. A family of universal curves yielding the capacitance of specifically dimensioned cocentric annular-plate capacitors

as obtained by quite intricate analysis. Thus, Kobayashi's[24] solution, the only one available, for coplanar disks enfolds a very complicated analysis based on use of double hypergeometric series.

At the conclusion of his solution Kobayashi mentions the desirability of determining the capacitance of a capacitor consisting of coplanar circular-annular plates, remarks on the corresponding greater difficulties that would attend solution, and conjectures that this problem is intractable to rigorous analytic solution. Such is also the recent opinion of Aikawa. Lacking an equation enabling calculation of the capacitance of a cocentric annular-plate capacitor, Aikawa, Sato, and Ataka determined the several values of the capacitance needed in their investigation by analogous measurements effected in an electrolytic tank.[25] The capacitance of this biplate capacitor can, however, be obtained analytically to any desired degree of precision and over any desired range of geometry by the method of subareas developed earlier by the present authors in determining the capacitances of several single-plate capacitors.[26–29]

Accordingly, the prime purpose of this paper is to advance, as determined by the method of subareas, a set of curves which afford the capacitance of a coplanar annular-plate capacitor of specified dimensions over the range of geometry to be encountered in practice.

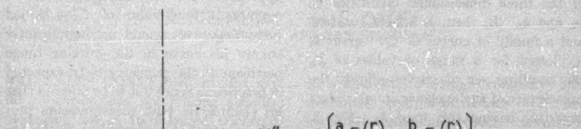

$V''=4GR\left\{\frac{a}{r}E\left(\frac{r}{a}\right)-\frac{b}{r}E\left(\frac{r}{b}\right)\right\}$
b
a
r''
r
r'
$V=4GR\left\{\frac{a}{r}E\left(\frac{r}{a}\right)-E\left(\frac{b}{r}\right)+k_b^2K\left(\frac{b}{r}\right)\right\}$
$V'=4GR\left\{E\left(\frac{a}{r}\right)-E\left(\frac{b}{r}\right)-k_a^2K\left(\frac{a}{r}\right)+k_b^2K\left(\frac{b}{r}\right)\right\}$
$k_a^2=1-\frac{a^2}{r^2};\ k_b^2=1-\frac{b^2}{r^2}$

Fig. 2. Potentials at a point r in the plane of a uniformly charged subannulus

Recasting the Computations

The basic course of procedure, somewhat different from that used in earlier papers,[26–29] is summarized as follows. Consider each of the two plates as comprised of subannuli, numbered $(i=1, \ldots, n)$ for one plate and $(i=n+1, \ldots, n+n')$ for the other. As in Fig. 1 therefore $n=6$ and $n'=4$. In view of angular symmetry, the charge density is a function of only the radial co-ordinate. Accordingly, consider each subannulus S_i as of essentially uniform, but different, charge density q_i and of essentially constant potential V_i. If V_{1i} denotes the potential at the midpoint of subannulus S_1 produced by the subannulus S_i of charge density q_i, the total potential V_1 of subannulus S_1 of charge density q_1 is

$$V_1=\sum_{i=1}^{n+n'} V_{1i} \qquad (1)$$

A similar equation expresses the potential

Table I. Capacitance as a Function of k_1r and k_2r ($k_3r=6$)

k_1r	k_2r	C/r, μμF	Number of Subannuli Used
1	6	0	
2	2	2.52	8
2.5	2.5	3.43	10
2.5	3	2.63	10
2.5	4	2.15	8
3	3	4.28	10
3	4	2.86	8
3.5	3.5	5.11	10
3.5	4	3.86	10
4	4	5.88	10
4	5	3.73	10
4.5	4.5	6.62	10
5	5	7.42	12
5.5	5.5	7.99	12
5.75	5.75	7.21	12
6	6	0	

of each of the other subannuli. Hence the resulting set of $n+n'$ equations is to be expressed generally by

$$V_j=\sum_{i=1}^{n+n'} V_{ji} \quad (j=1,\ldots,n+n') \tag{2}$$

where each of the V_{ji} are furnished by appropriate use of the equations of Fig. 2, which have been developed in reference 27, equations 7, 8, 9.

Imposing next the conditions, stemming from the fact that in exact solution each plate is an equipotential surface, $V_0=V_j(j=1,\ \ldots,n)$ and $V_0'=V_j(j=n+1,\ldots,n+n')$, and subtracting in turn each equation (furnished by $j=2,\ \ldots,n+n'$) from the first yields a set of $(n+n'-1)$ equations for the q_i. Of this the constant member on the right-hand side of each equation is either zero or (V_0-V_0'). An additional equation stems from the condition that in exact solution the plates have equal but opposite charges, hence the sum of the charges on the $n+n'$ subannuli is zero. Thus,

$$\sum_{i=1}^{n+n'} Q_i=\sum_{i=1}^{n+n'} A_iq_i=0 \tag{3}$$

A_i designates the area of the subannulus S_i. Solution of the complete set of $(n+n')$ equations 2 and 3 yields expression of the q_i in terms of (V_0-V_0').

Finally, as in previous solutions,[26-29] approximation of the value of the capacitance C follows from

$$C=Q/(V_0-V_0')=\left[\sum_{i=1}^{n} A_iq_i\right]\Big/(V_0-V_0') \tag{4}$$

The term (V_0-V_0') cancels on effecting the division indicated in equation 4. In addition, the values of the q_i afford approximation to the charge density distribution over the plates.

In conclusion, it is to be noted that determination of C and $q(r)$ can be effected to increasingly good degrees of precision by an increase of n and n'. As indicated in the following section, excellent approximations are afforded by use of relatively small values of n and n'.

An Illustrative Example

Fig. 1 shows one quadrant of a specifically dimensioned annular capacitor (of radii 1 and 4 cm (centimeter) and 5 and 6 cm) divided into 10 subannuli. Evaluating the V_{ji} in equation 2 by use of the equations tabulated in Fig. 2 and the Q_i in equation 3 yields a set of 10 equations in terms of the 10 unknown q_i. Solution yields values of q_i as noted on Fig. 1. Hence

$$Q_{\text{total}}=\sum_{i=6}^{10} A_iq_i=3.353(V_0-V_0') \tag{5}$$

Equations 4 and 5 furnish $C=Q/(V_0-V_0')=3.353(V_0-V_0')/(V_0-V_0')=3.353$ electrostatic units $=3.726$ μμf (micromicrofarad), as plotted at the point ($k_1=4$, $k_2=5$) on Fig. 3.

Universal Curves for Capacitance

A series of computations, effected as in the preceding section, where the values of $r=1$ cm and $k_3=6$ cm are maintained and k_1 and k_2 are suitably varied between 1 and 6, yields the data of Table I.

As in the electrostatic units set of units, the primary dimensions of C are centimeters; and as in virtue of angular symmetry the capacitance of a coplanar annular-plate capacitor is a function of the four radial dimensions alone, it follows from Buckingham's Π-Theorem[30] that C/r is dimensionless and is a function of the three dimensioned quantities k_1, k_2, and k_3. If, then, k_3 is held constant and a family of curves of C/r versus k_1 is plotted for a range of values of k_2, the resulting set of curves affords the capacitance of any coplanar annular-plate capacitor having this fixed k_3. Fig. 3, plotted from the data of Table I, comprises this family of universal curves. The value of $r=1$ shown on Fig. 1 is one of convenience in the solution for the normalized curves of Fig. 3; any arbitrary value r_0 of r could be chosen. Thus, taking $r_4=6r_0$, whence k_3 yet equals 6, yields the same set of C/r curves.

The values of $C/r=0$ at the points ($k_1=1$, $1\leq k_2\leq 6$) and the corresponding values of $C/r=0$ at the points ($1\leq k_1\leq 6$, $k_2=6$) are for points where the inner and outer annuli, respectively, degenerate to circular lines, for which $C=0$: which is always the case for a capacitor comprised of two "plates" where one plate is a finite surface and the other plate degenerates to a finite line segment.

Finally, a plot of a similar set of curves for a desired range of values for k_3 would yield a portfolio of universal curve sheets from which could be obtained, by means of suitable interpolation, the capacitance of any specifically dimensioned capacitor.

Experimental Confirmation

For the annular capacitor defined by ($r=0$, $k_1=5/2$, $k_2=3$, $k_3=6$) Aikawa found (Fig. 5 of reference 3) by electrolytic tank measurements that $C\cong 2.75$ μμf. For the same k_i ($i=1, 2, 3$) but $r=1.0$, Fig. 3 yields, in good agreement, $C\cong 2.63$ μμf. The slightly higher value given by Aikawa is to be expected: for his inner plate encompasses a complete circular disk ($r=0$) rather than an annulus ($r=1$). The center holes in the configuration of the annuli act to reduce the capacitance slightly, how much depending upon the magnitude of k_1/r_1: as predicted by the well-known theorem in electrostatics that the capacitance of any portion of a capacitor is less than the capacitance of the complete capacitor.

It is of interest to note that Aikawa's electrolytic tank measurements tend to a final value of $C=\infty$ as the radial separation between the plates approaches zero (the edge-effect becoming the predominant factor); such results from the necessary use of plates of finite, even though very thin, thickness. However, such is not the case for the indefinitely thin plates of this paper. Thus in Fig. 3 for the limiting case of $k_1=k_2=3$, the capacitance approaches $C\cong 4.25$ μμf, whereas Aikawa's "corresponding" value approaches $C=\infty$. Changing Aikawa's curve to a final value of $C\cong 4.25$ μμf (assuming $r=0$) would not significantly change his curve in the working range pertinent to the geometry to be expected in practice: whence it is to be inferred that his electrolytic tank measurements provide substantial confirmation of the accuracy of the calculated curves of Fig. 3.

Summary

The basic theory of determination of the capacitance of a coplanar annular-plate capacitor by the method of subareas (subannuli) is outlined and application is illustrated by detail of calculation for a specifically dimensioned capacitor. Corroboratively, this value is found to be in good agreement with corresponding experimental values determined by electrolytic tank measurement.

A family of universal curves is advanced, which yields the capacitances of annular capacitors over the range of radial geometry of essential interest in practice.

References

1. On the Capacity of Concentric Ring Electrodes (in Japanese), **K. Aikawa.** *Paper no. 24A1*, Convention, Tohoku Branch, Institute of Electrical Engineers of Japan, Tokyo, Japan, Oct. 1957.

2. On the Capacity of Concentric Ring Electrodes (in Japanese), **T. Oshiyama, K. Aikawa, T. Sano, M. Nakaya.** *Report*, Faculty of Engineering, Yamanashi University, Yamanashi, Japan, no. 6, 1955, p. 135.

3. On the Capacity of Concentric Ring Electrodes—Second Report (in Japanese), **T. Oshiyama, K. Aikawa, T. Sano.** *Report*, Faculty of Engineering, Yamanashi University, no. 7, 1956, p. 47.

4. The Capacity Between a Circular Plate and a Concentric Outer Ring Plate (in Japanese), **H. Ataka.** *Technical Papers*, Institute of Electrical Engineers of Japan, vol. 4, 1943, p. 185.

5. (In Japanese), **S. Sato.** *Journal*, Institute of Electrical Engineers of Japan, vol. 64, 1944, p. 75.

6. (In Japanese), **H. Ataka.** *Technical Papers*, Institute of Electrical Engineers of Japan, vol. 5, 1944, p. 9.

7. Problems Relating to a Thin Plane Annulus, **J. W. Nicholson.** *Proceedings*, Royal Society of London, London, England, vol. 101A, 1922, pp. 195–210.

8. The Flow of Viscous Fluids Round Plane Obstacles, **R. Roscoe.** *Philosophical Magazine*, London, England, ser. 7, vol. 60, 1949, pp. 338–51.

9. The Forces on a Flat Plate Placed in a Stream of Fluid Between Two Parallel Walls, **S. Tomotika.** *Proceedings*, Physico-Mathematical Society of Japan, Tokyo, Japan, ser. 3, vol. 14, 1932, pp. 139–67.

10. A Treatise on Electricity and Magnetism (book), **J. C. Maxwell.** Oxford University Press, London, England, vol. 1, third edition, 1893, p. 239.

11. Electrical Researches (book), **H. Cavendish, J. C. Maxwell,** editor. Cambridge University Press, Cambridge, England, 1879, pp. 394, 396, 424.

12. Ueber die Anordnung der Elektricitat auf Einer Einzelnen Sehr Dünnen Platte und auf den Beiden Belegungen Einer Franklin'schen Tafel, **B. Clausius.** *Annalen der Physik und Chemie*, Leipzig, Germany, vol. 86, 1852, pp. 161–206.

13. Zur Theories des Condensators, **G. Kirchhoff.** *Monatsberichte*, Akademie der Wissenschaften, Berlin, Germany, 1877; also Gesammelte Abhandlungen, Leipzig, Germany, vol. 1, 1882, pp. 101–17.

14. Kreisscheibenkondensator, **W. V. Ignatowski.** *Comptus Rendus*, l'Académie des Sciences de l'USSR, Leningrad, U.S.S.R., Travaux de l'Institute Stekloff, II, vol. 3, 1932, p. 104.

15. Isoperimetric Inequalities in Mathematical Physics (book), **G. Polya, G. Szegö.** Princeton University Press, Princeton, N. J., 1951, p. 78.

16. The Electrification of Two Parallel Circular Discs, **J. W. Nicholson.** *Transactions*, Royal Society of London, vol. 224A, 1924, pp. 303–69.

17. The Electrostatic Problems of Two Equal Parallel Circular Plates, **Y. Nomura.** *Proceedings*, Physico-Mathematical Society of Japan, vol. 23, 1941, pp. 168–79.

18. The Electrostatic Field of Two Equal Circular Coaxial Conducting Discs, **E. R. Love.** *Quarterly Journal of Mechanics and Applied Mathematics*, London, England, vol. 2, 1949, pp. 428–51.

19. Numerical Calculation of Certain Small Electrostatic Effects, **R. Cade, K. A. Small.** *Journal of Applied Physics*, New York, N. Y., vol. 29, no. 1, 1958, pp. 53–5.

20. A Solution of Tranter's Dual Integral Equations Problem, **J. C. Cooke.** *Quarterly Journal of Mechanics and Applied Mathematics*, vol. 9, 1956, pp. 103–10.

21. The Coaxial Circular Disc Problem, **J. C. Cooke.** *Zeitschrift fur Angewandte Mathematik und Mechanik*, Berlin, Germany, vol. 38, no. 9/10, 1958, pp. 349–56.

22. The Approximate Solution of Dual Integral Equations by Variational Methods, **B. Noble.** *Proceedings*, Edinburgh Mathematical Society, Edinburgh, Scotland, vol. II, pt. 2, 1958, pp. 115–26.

23. Random Walk Related to the Capacitance of the Circular-Plate Condenser, **E. Reich.** *Quarterly of Applied Mathematics*, Providence, R. I., vol. 11, 1953, pp. 341–45.

24. Das Elektrostatische Potential um Zwei auf Derselben Ebene Liegende und Sich Nicht Schneidende Gleichgrosse Kreisscheiben, **I. Kobayashi.** *Science Reports*, Tohoku Imperial University, Sendai, Japan, vol. 27, 1939, pp. 365–91.

25. Experimental Determination of the Capacitance of Heavy-Current Busses Comprised of Solid or Tubular Rectangular Conductors, **T. J. Higgins, R. D. Teasdale.** *AIEE Transactions*, vol. 67, pt. I, 1948, pp. 653–59.

26. Calculation of the Electrical Capacitance of a Cube, **D. K. Reitan, T. J. Higgins.** *Journal of Applied Physics*, vol. 22, 1951, pp. 223–26.

27. Calculation of the Capacitance of a Circular Annulus by the Method of Subareas, **T. J. Higgins, D. K. Reitan.** *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 926–33.

28. Accurate Determination of the Capacitance of a Thin Rectangular Plate, **D. K. Reitan, T. J. Higgins.** *Ibid.*, pt. I (*Communication and Electronics*), vol. 75, 1956 (Jan. 1957 section), pp. 761–66.

29. Accurate Determination of the Capacitance of Rectangular Parallel-Plate Capacitors, **D. K. Reitan.** *Journal of Applied Physics*, vol. 30, 1959, pp. 172–76.

30. Mathematics of Modern Engineering (book), **R. E. Doherty, E. G. Keller.** John Wiley & Sons, Inc., New York, N. Y., 1936, p. 144.

Detection of Transmission Errors in 5-Level Punched Tape

R. STEENECK
NONMEMBER AIEE

H. F. CALEY
ASSOCIATE MEMBER AIEE

BUSINESS CONCERNS in this age of automation are making steadily increasing use of their leased telegraph systems for the transmission of data in 5-level tape form. The number of errors introduced in the data by transmission over high-grade facilities is relatively small and can be tolerated in many business applications. In some situations, however, the error frequency must be reduced to the lowest possible rate. For these situations the telegraph company has developed equipment which will incorporate redundant check characters in the data at the transmitting end of the circuit and perform the check and delete the check characters at the receiving end.

Error-Checking Principle

The first problem which must be resolved in the design of an error-detecting system is which of the many possible error-detecting codes will be used. As this system had to be compatible with conventional 5-level telegraph switching equipment, use of any of the simple vertical parity checks was ruled out because they require at least six levels. In addition, extensive tests showed that vertical parity checks are relatively ineffective when the bits making up a character are transmitted serially as they are on a telegraph circuit. Approximately 10% of the errors encountered in telegraphic communications affect two bits in the same character and the resulting errors are not detected by the simple vertical odd or even parity check. The information which can be transmitted in any given time is also reduced by $16^{2}/_{3}\%$. More complex vertical parity checks have been devised which lower the 10% undetected error rate but they further decrease the rate of transmission.

Attention was therefore directed to the horizontal-type checks and tests were made of the effectiveness of a single horizontal parity character at the end of each line (maximum of 72 characters) and a two-character binary total at the end of each line. Both of these checks proved superior to the vertical parity checks with the two-character binary total being the more effective of the two. It had the further advantage of reducing transmitted information per unit time by less

---

Paper 59-1147, recommended by the AIEE Data Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted June 12, 1959; made available for printing August 4, 1959.

R. Steeneck and H. F. Caley are with the Western Union Telegraph Company, New York, N. Y.

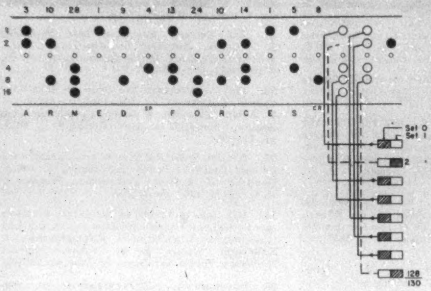


Fig. 1. Preparation of tape at sending stations

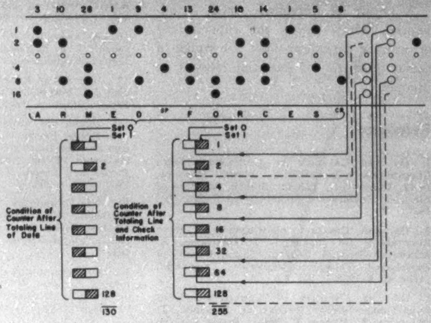

3 10 28 1 9 4 13 24 10 14 1 5 8
1
2
4
8
16
A R M E D SP F O R C E S CR
Set 0
Set 1
2
128
130
Condition of Counter After Totaling Line of Data
Condition of Counter After Totaling Line and Check Information
Set 0
Set 1
1
2
4
8
16
32
64
128
255

Fig. 2. Checking at receiving station

than 5%. Accordingly, this system was adopted.

Figs. 1 and 2 illustrate the basic features of the 2-character binary total-error detection system. Each character is considered a 5-digit binary number with levels 1–5 having weights 1, 2, 4, 8, and 16 respectively. The characters in a variable-length line which does not exceed 72 characters are totalled up to and including the "carriage return."

The simple computer used has 8 stages and can therefore total from 0 to 255. As the total for a line of 72 characters can be as high as 2,232 the counter will cycle a number of times in totaling an average line. This raises the possibility that equipment troubles in switching centers such as sticking pins on transmitters will consistently add or drop a number of pulses in one of the levels which may be sufficient to exactly cycle the counter and cause the errors to be undetected. This cannot occur in levels 1 and 2 when block length is limited to 72 characters because 256 pulses must be lost or gained in level 1 and 128 pulses lost or gained in level 2. However, the possibility does exist in levels 3, 4, and 5 where gaining or losing 64, 32, and 16 pulses, respectively, will cycle the counter.

A simple modification which causes the computer to cycle on a count of 255 instead of 256 greatly improves this situation and requires that 255 pulses be added or lost in anyone of the 5 levels before a computer cycle can be gained or lost. This, of course, cannot occur in 72-character groups.

At the end of the group the complement of the 8-digit binary total in the computer is punched in the tape as levels 1, 2, 4, and 5 in two characters. The third level of each of these characters is always punched to avoid the formation of control-character combinations such as *Figs*, *H* and *CR*, *CR* used in switching systems.

The checking process at the receiving station consists of again totaling the characters in the group in a computer very similar to the one in the sending unit. The 8 binary digits in the two check characters are also totalled and if no errors were introduced in transmission this results in all eight stages of the computer being set on 1. The binary additions in Table I illustrate the processes taking place in the sending and receiving computers for the short blocks of data shown on Figs. 1 and 2.

If any of the bits in the block were reversed during transmission it can be seen that the foregoing process would result in at least one of the stages in the receiving computer being set on zero.

With this checking system errors occurring in one character can be compensated for by equal and opposite errors in the same level in another character in the same block of data. This condition, fortunately, is quite rare on high-grade land lines. Much more frequent is the mutilation of several successive bits by one "hit" on the line or the shifting of a marking bit from one level to the next and this system prevents such errors from being compensating.

System Design

Both sending and receiving units operate at 200 wpm (words per minute) and are, therefore, capable of handling up to three 65-wpm transmission circuits.

The equipment can be packaged in different ways depending on whether data will flow in only one direction or in both directions. The simplest system for checking data flowing in one direction will be described but it will be noticed that sending and receiving station equipments are very similar and with only minor additions and modifications could serve both functions.

The checking equipment at the sending station is illustrated in Fig. 3 and a block diagram is shown in Fig. 4.

Table I. Computations at Sending and Receiving Stations

	Sending Station		Receiving Station	
Character	Binary Number	Decimal Number	Binary Number	Decimal Number
A	00011	3	00011	3
R	01010	10	01010	10
M	11100	28	11100	28
E	00001	1	00001	1
D	01001	9	01001	9
Space	00100	4	00100	4
F	01101	13	01101	13
O	11000	24	11000	24
R	01010	10	01010	10
C	01110	14	01110	14
E	00001	1	00001	1
S	00101	5	00101	5
CR	01000	8	01000	8
Total	10000010	130	10000010	130
Complement	01111101	125	Complement 01111101	125
			Total 11111111	255

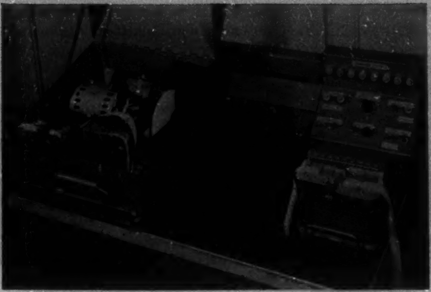


Fig. 3. Sending station equipment

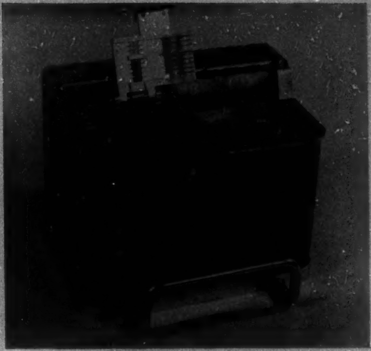


Fig. 5. Dual-reading tape transmitter

Sending Station Operation

The sending-end equipment is primarily designed to produce from the data tape a new tape incorporating check characters at the end of each line of data. To start this process the selector switch is set to the "Generate Tape" position and the data tape without check characters which has been prepared on a *Type 19* set, or other tape producing equipment, is fed through the special dual-reading transmitter; see Fig. 5. This transmitter has two sets of reading pins 0.1 inch apart which is the distance between characters on the tape. The first set feeds the computer and the second set feeds the reperforator through the check character switch and the character-reading matrix. Routing information at the beginning of the tape for the control of switching centers is transmitted to the reperforator, but is not sent to the computer which is turned off at this time. At the end of the routing information, a special 3-character code which has been inserted in data preparation is read indicating the beginning of the data. At this point the computer is turned on automatically and set to zero count. In the block diagram the starting 3-character code is shown as *Figs, A, Ltrs;* however, any combination could be used.

The entire first block of data ending with *CR* is sent to both the reperforator and the computer. When *CR* is read by the matrix the transmitter is stopped, the two check characters are transmitted from the computer to the reperforator and the computer is reset to zero. This process is repeated for the next data block beginning with *LF* and ending with *CR* and for succeeding data blocks until the *end-of-data* characters *Figs, A, F, Ltrs* are read. This combination turns off the computer and allows the message terminating characters to be transmitted only to the reperforator. By using two sets of reading pins in the transmitter, one to transmit to the computer and one to the reperforator, the possibility of reading errors being undetected is practically eliminated. The possibility of both sets of pins making the same error is extremely small and any difference will be detected in the check at the receiving station.

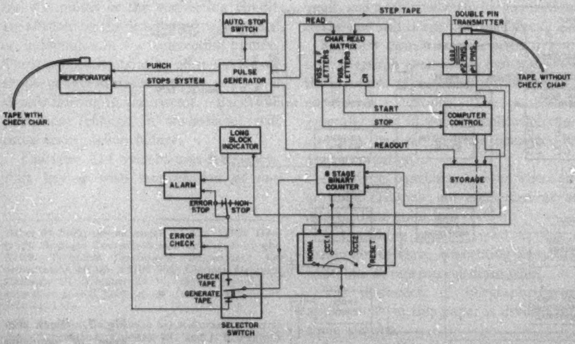

AUTO. STOP SWITCH
STEP TAPE
READ
DOUBLE PIN TRANSMITTER
CHAR. READ MATRIX
FIGS. A, F LETTERS
FIGS. A LETTERS
CR
#2 PINS
#1 PINS
TAPE WITHOUT CHECK CHAR
REPERFORATOR
PUNCH
STOPS SYSTEM
PULSE GENERATOR
TAPE WITH CHECK CHAR.
START
STOP
COMPUTER CONTROL
LONG BLOCK INDICATOR
READOUT
ALARM
8 STAGE BINARY COUNTER
STORAGE
ERROR STOP
NON-STOP
ERROR CHECK
NORM.
CCT 1
CCT 2
RESET
CHECK TAPE
GENERATE TAPE
SELECTOR SWITCH

Fig. 4. Block diagram of sending station

The tape that has been produced in the reperforator contains check characters for each block of data and may be transmitted over any conventional switching system or leased circuit.

If it is desired to check the tape locally transmitting it the selector switch is thrown to "Check Tape" which disconnects the reperforator and connects checking units. The tape is fed through the dual-pin transmitter and each block is checked for punching errors in a manner which will be described in succeeding sections covering the receiving station unit. If an error is detected an alarm is sounded and the tape stops until the "Reset" switch is operated which silences the alarm and causes checking to continue.

Receiving

At the receiving station the data tape with check characters is received on a reperforator and is fed to the transmitter on the error-checking unit, shown in Fig. 6. A block diagram of this receiving unit indicating the functions of the various components is shown in Fig. 7.

The transmitter sends to a reperforator through the reperforator control switching unit which is in the normal position at this time. Read-back pins on the reperforator spaced 5 characters after the punch pins read the final tape and feed the characters to the computer and the character-reading matrix. The

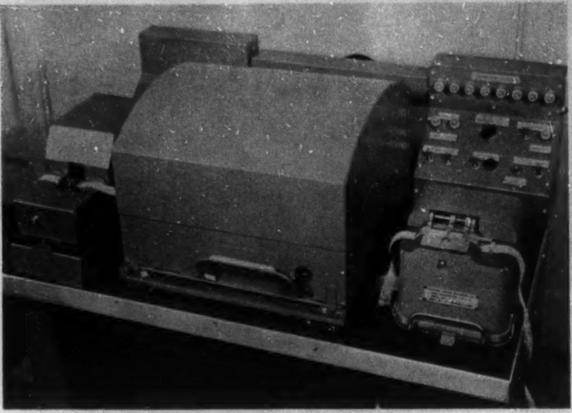


Fig. 6. Receiving station equipment

computer is turned off while the routing information at the beginning of the tape is passing over the read-back pins. When the characters *Figs*, *A*, *Letters* are read by the matrix associated with the reperforator the computer is turned on and totaling of the characters in the first data block starts.

The checking process is done in two different ways depending on whether the selector switch is in "Delete" or "Non-Delete" position. If it is desired that check characters shall not be punched in the final tape the Selector Switch is thrown to "Delete" and checking proceeds in the following manner.

The first data block continues to be sent from the transmitter to the reperforator until the matrix associated with the transmitter reads the *CR* at the end of the block. It should be remembered that the read-back pins on the reperforator are reading the 5th character before the *CR* because of the spacing of the punch pins and the read-back pins. The reperforator is turned off at this point and levels 1, 2, 4, and 5 of the two check characters are transmitted only to the computer. The computer control unit directs the four elements of the first character to stages 1–4 of the computer and the four elements of the second check character to stages 5–8 of the computer. The reperforator is then turned on again and the transmitter starts to send the characters of the second data block starting with *LF* and the read-back pins on the reperforator read the remaining characters of the first data block. When the read-back pins read the *CR* at the end of the first block this is detected by the character reading matrix associated with the reperforator. After the *CR* character is totaled a check is made to see if all stages of the counter are set to 1 and the counter is reset. If this check is satisfied the block is marked in black by the tape marker and checking of the second block proceeds without interruption. If any counter stage is set to 0 the system is stopped, an alarm is sounded and the errored block is marked in red. The "Reset" switch must be operated before checking will proceed. However, if the error switch is thrown to "Non-Stop" detection of an error will only cause red marking of the block. Checking of succeeding blocks will be continued and no alarm will be sounded.

This checking process proceeds until the reading matrix associated with the reperforator detects *Figs*, *A*, *F*, *Letters* after the check characters at the end of the last data block. The computer is turned off and any further characters in the received tape are punched but not checked.

To safeguard against improper operation of the checking circuitry a long-block indicator is provided to detect the absence of a check cycle within the maximum block length of 72 characters. This condition stops the equipment and sounds an alarm indicating that an end-of-block signal and a check cycle have been missed. Action can then be taken to recheck and if necessary rerun the skipped block.

Because a low error rate is anticipated

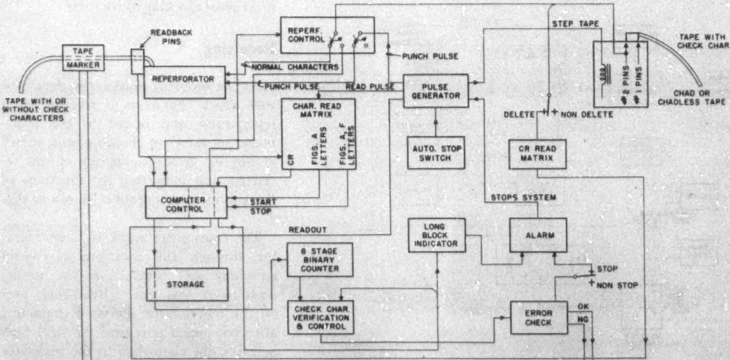

READBACK PINS
TAPE MARKER
TAPE WITH OR WITHOUT CHECK CHARACTERS
REPERF. CONTROL
REPERFORATOR
NORMAL CHARACTERS
PUNCH PULSE
PUNCH PULSE
READ PULSE
PULSE GENERATOR
STEP TAPE
TAPE WITH CHECK CHAR.
# 2 PINS
# 1 PINS
CHAD OR CHADLESS TAPE
DELETE
NON DELETE
CHAR. READ MATRIX
CR
FIGS. A LETTERS
FIGS. A, F LETTERS
AUTO. STOP SWITCH
CR READ MATRIX
COMPUTER CONTROL
START
STOP
STOPS SYSTEM
READOUT
LONG BLOCK INDICATOR
ALARM
8 STAGE BINARY COUNTER
STORAGE
STOP
NON STOP
CHECK CHAR. VERIFICATION & CONTROL
ERROR CHECK
OK
NG

Fig. 7. Block diagram of receiving station

on normal facilities and an inexpensive system was desired automatic error correction is not provided. When an error is detected the block must be identified and a rerun requested from the sending station.

Conclusions

At the present time only empirical figures can be given on the effectiveness of this system in detecting transmission errors. A calculated figure based on an assumption of a random distribution of single-bit errors throughout the levels and blocks would be meaningless as errors introduced in switching equipment have a different distribution from those introduced on the lines, and neither appear to be randomly distributed single-bit errors. A mathematical analysis of effectiveness will have to await the results of tests now under way to determine the rate, nature, and distribution of errors introduced by lines and equipment. However, while the precise distribution of errors is not yet known, experience and tests to date have given an indication of the general pattern of equipment and line errors.

This system has been designed to be particularly effective in detecting both of these types of errors while at the same time requiring a low degree of redundancy. Operational tests on typical switching systems and frequency-modulated telegraph circuits indicate that when the telegraph system error rate is one errored character per 50,000 characters transmitted, the undetected error rate will be about one undetected error per 100,000,000 characters transmitted.

Application of a New Carrier Telegraph System

T. B. COLLINS, JR.
ASSOCIATE MEMBER AIEE

J. E. PITTS
NONMEMBER AIEE

THE transmission of telegraph or data signals by carrier frequencies operating in the voice-frequency band involves considerations not normally associated with voice signals transmitted in this band. Transmission requirements for the data or telegraph carrier facility may range from those of a single-channel 2-wire system to the maximum of a 26-channel 4-wire system.

The bandwidth of the carrier facility is a determining factor in the number of channels that may be carried by the circuit. The signal-to-noise ratio has a direct bearing on the amount of distortion introduced into the telegraph signals.

When a telegraph channel is remote from the data or teleprinter equipment, the d-c pulses in the connecting circuit are affected by the distributed capacitance or inductance in the connecting circuit. These properties tend to alter the waveshape, or rate of rise and decay, of the d-c signals flowing in the circuit. Excessive distortion (change in waveshape) will cause transmission failure.

The type *23A* system uses frequency-shift keying with transmission of one frequency indicating a mark signal, and transmission of another frequency indicating a space signal. A block diagram of a typical channel with complementary equipment panels is shown in Fig. 1. A complete operating description of the system is included in a paper by Stephens.[1]

Fig. 2 shows a channel shelf with one unit disengaged from its mounting.

A subscriber set, Fig. 3, is used to install a single-channel unit at a subscriber's premises. The subscriber set consists of a cabinet to house a channel unit and the local d-c power supply. Contained in the cabinet as part of the set are 2- or 4-wire line-terminating equipment for the channel unit, loop-terminating equipment, and a terminal strip for connections to the line and teleprinter machine.

Auxiliary panels may be used to connect telegraph channels in multiple when the channels of a system are terminated at different locations. The panels are generally used in 2-wire applications but also may be used in some instances for 4-wire operation

A bridging panel may be used where one or more channels are connected at an intermediate point in a system.

A branching panel is used to join a maximum of three connecting lines to a main line at one intermediate point.

The discussion of telegraph carrier applications in this paper is divided into three sections.

1. Transmission considerations.
2. System application considerations.
3. Direct-current loop arrangements.

The equipment referred to in this paper is a new system developed by the Lenkurt Electric Company, the type *23A* Telegraph and Data Carrier System.

Transmission Considerations

Channel Frequencies

The frequency assignments of the *23A* system are shown in Tables I and II. These are not the actual frequencies transmitted; the transmitted frequencies are obtained by adding and subtracting the amount of specified shift from the mid-band frequency.

Low-Frequency Band Channels

The first 18 channels are generally considered as the low-frequency group. The frequency band of this group falls between 340 and 3,400 cps (cycles per second). The only requirement of frequency response over this band is that the difference in level between the mark and space frequency of any given channel should not exceed 1 db (decibel) for minimum effect on received signals bias. A greater difference is not detrimental in itself, but usually indicates a drop-off in filter response (low-pass filters, band separation filters, channel filters) which may cause a change in the velocity of transmission between the mark and space frequencies. This may, in turn, result in bias of the d-c signal output of the discriminator.

A greater slope than 1 db can, of course, be tolerated and the resulting bias can be removed by adjustment of the bias potentiometer during system lineup. However, if the channel is patched to a spare facility with a different slope characteristic, a change in the bias adjustment would probably be required for optimum performance.

Paper 59-1107, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11-16, 1959. Manuscript submitted June 8, 1959; made available for printing July 14, 1959.

T. B. Collins, Jr., is with the Automatic Electric Sales Corporation, Chicago, Ill., and J. E. Pitts is with Lenkurt Electric Company, Inc., San Carlos, Calif.

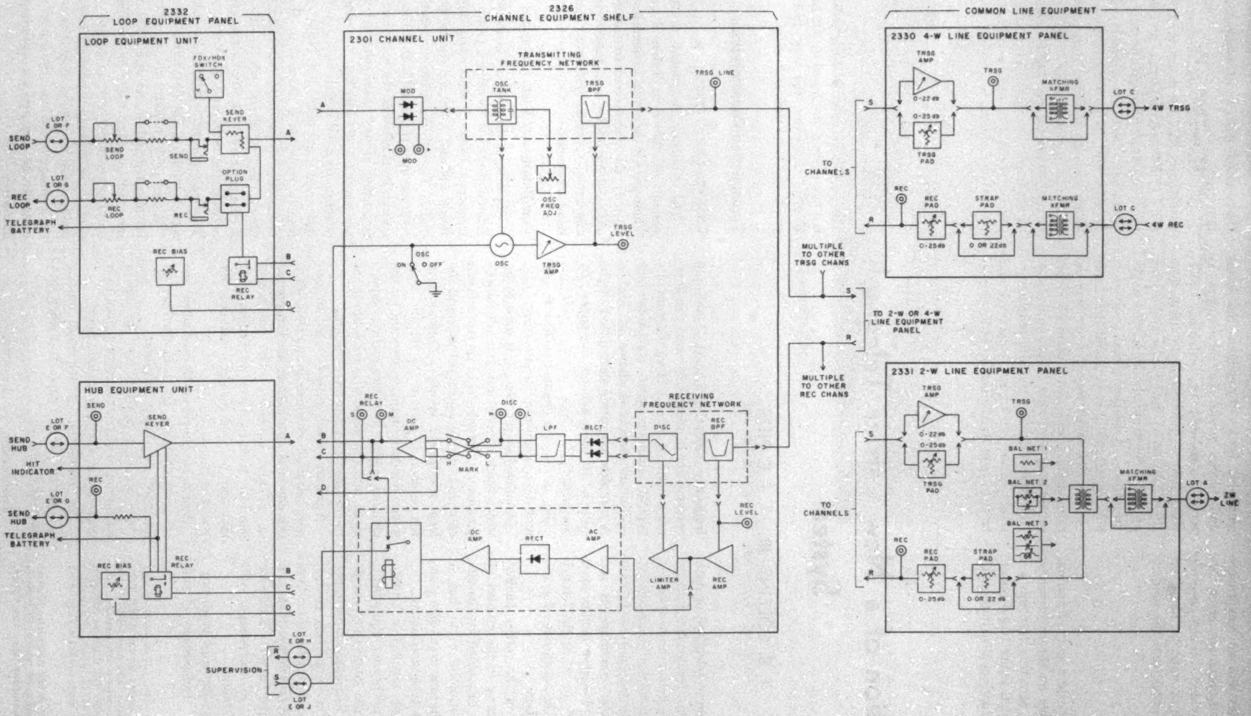

2332
LOOP EQUIPMENT PANEL
LOOP EQUIPMENT UNIT
FDX/HDX SWITCH
SEND KEYER
SEND LOOP
LOT E OR F
SEND LOOP
SEND
A
REC LOOP
LOT E OR G
REC LOOP
REC
OPTION PLUG
TELEGRAPH BATTERY
REC BIAS
REC RELAY
B
C
D
HUB EQUIPMENT UNIT
SEND
SEND KEYER
SEND HUB
LOT E OR F
HIT INDICATOR
REC
SEND HUB
LOT E OR G
TELEGRAPH BATTERY
REC BIAS
REC RELAY
SUPERVISION
R
S
LOT E OR H
LOT E OR J
2326
CHANNEL EQUIPMENT SHELF
2301 CHANNEL UNIT
TRANSMITTING FREQUENCY NETWORK
MOD
OSC TANK
TRSG BPF
TRSG LINE
OSC FREQ ADJ
OSC
ON
OFF
TRSG AMP
TRSG LEVEL
REC RELAY
S
M
DC AMP
MARK
H
L
DISC
LPF
RECT
RECEIVING FREQUENCY NETWORK
DISC
REC BPF
REC LEVEL
DC AMP
RECT
AC AMP
LIMITER AMP
REC AMP
TO CHANNELS
MULTIPLE TO OTHER TRSG CHANS
TO 2-W OR 4-W LINE EQUIPMENT PANEL
MULTIPLE TO OTHER REC CHANS
COMMON LINE EQUIPMENT
2330 4-W LINE EQUIPMENT PANEL
TRSG AMP
0-22db
0-25db
TRSG PAD
TRSG
MATCHING XFMR
LOT C
4W TRSG
REC
REC PAD
0-25db
STRAP PAD
0 OR 22db
MATCHING XFMR
LOT C
4W REC
2331 2-W LINE EQUIPMENT PANEL
TRSG AMP
0-22db
0-25db
TRSG PAD
TRSG
BAL NET 1
BAL NET 2
BAL NET 3
MATCHING XFMR
LOT A
2W LINE
REC
REC PAD
0-25db
STRAP PAD
0 OR 22 db
S
R

Fig. 1. Block diagram of 23A system

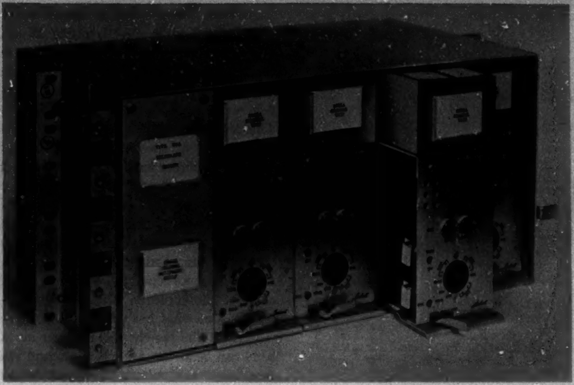


Fig. 2. Channel equipment shelf with four channel units, showing one unit partially removed

When the connecting facility is nonloaded cable, which has a gradually rising attenuation characteristic at the higher frequencies, the only limitation other than the 1-db restriction mentioned will be the attenuation offered to the frequency of the highest channel to be transmitted.

HIGH-FREQUENCY BAND CHANNELS

The eight channels above 3,500 cycles are referred to as the high-frequency group. The same frequency response considerations will generally apply in the high-frequency group as outlined for the low-frequency channels. Where low-frequency telephone carrier is operated on the same facility as the high-frequency group, the band-separation filter cut-off frequency will be the determining factor. When the frequency assignment of an high-frequency telegraph channel falls near the knee of the band-separation filter, the difference in response between the mark and space frequencies should not exceed 1 db for minimum effect on telegraph signal bias.

Table I. Frequency Assignments, 23A System, for Low-Frequency Group

Channel Spacing 170 Cycles; Frequency Shift ±35 Cycles

Mid-Band Frequency, Cps	Mid-Band Frequency, Cps
425	1,955
595	2,125
765	2,295
935	2,465
1,105	2,635
1,275	2,805
1,445	2,975
1,615	3,145
1,785	3,315

DISTORTION

Distortion in telegraph and data signals can come from several sources. The drop equipment (printer machines) and the connecting loop to the carrier terminal are, of course, heavy contributors to the total distortion in a telegraph or data circuit. In the carrier portion of the circuit the sources of distortion are the channel terminal circuitry and relays, and the connecting circuit between the two terminals. The *23A* channel equipment has been designed to hold total distortion to a minimum at speeds up to 100 words per minute.

Bias is a form of distortion in which the transmitted signals are uniformly shortened or lengthened. If marking signals received at the receiving terminal are longer than the signals transmitted at the sending end, they are said to have marking bias. If the received marking signals are shorter than those transmitted, the signals are said to have spacing bias.

Bias contributed by the carrier equipment is caused by frequency error, temperature variation, voltage changes, aging of components, and residual bias in the polar keying relay. All of these sources of bias can be easily compensated for in the initial line-up and subsequent routine tests of the *23A* system.

EFFECTS OF LEVEL CHANGES AND NOISE

Telegraph systems with frequency shift keying can tolerate more variation in circuit levels than can comparable systems with make-and-break or amplitude keying.

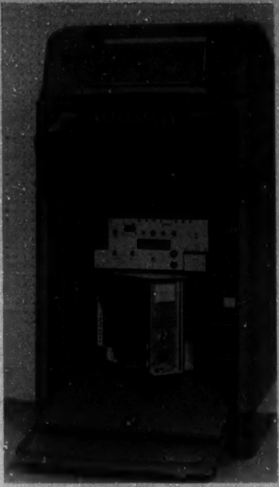


Fig. 3. Subscriber set mounted in model 28 teletypewriter console

In the *23A* system, with the channel discriminator preceded by a fast-acting limiter, input level changes of 35 db will produce less than 2% change in receiving bias.

Noise also has less effect on a frequency shift system than on an amplitude-modulated (AM) system. An AM system depends on the absence of tone (or carrier) as one of the indicating conditions. Noise present in the circuit during the tone-off period, if of sufficient amplitude, can cause errors in received data. In a frequency-modulated system, the tone (or carrier) is present for both indicating conditions; this tends to mask any noise present on the circuit.

SIGNAL LEVEL LIMITS

When telegraph carrier signals are transmitted over carrier-derived tele-

Table II. Frequency Assignments, 23A System, for High-Frequency Group

Mid-Band Frequency, Cps	Frequency Shift, Cps	Channel Spacing, Cps
3,550	±40	200
3,750	±40	200
3,950	±40	200
4,150	±40	210
4,360	±45	220
4,580	±45	230
4,810	±50	240
5,050	±50	

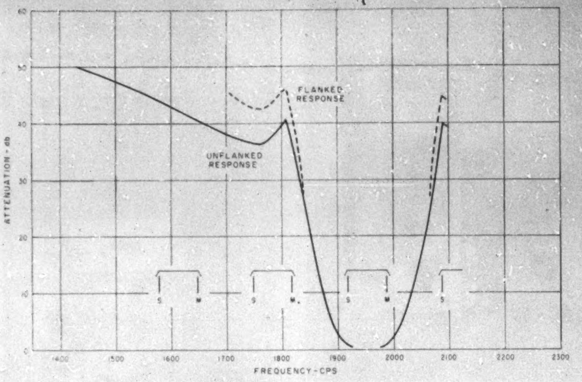


Fig. 4. Typical receive-filter-response low-frequency channels

phone channels, the transmit and receive levels are determined primarily by the requirements of the carrier telephone channel. The *23A* system has ample range to meet the requirements of any telephone carrier system. However, when *23A* systems are operated over physical circuits, the transmit and receive levels used will vary, depending on factors such as the type of circuit, the levels on adjacent circuits, and whether telephone carrier is also applied to the circuit.

The maximum total power output of either the single-channel or multichannel line amplifier in the *23A* system is +6 dbm (db referred to 1 milliwatt). Where crosstalk considerations and power-handling capability of the transmitting medium permit, the maximum possible line level should be used. This will improve the signal-to-noise ratio and increase the length of a physical circuit over which the *23A* system may be operated without repeaters.

As shown in Fig. 4, the minimum rejection of an unflanked filter to tones of an adjacent channel is slightly more than 35 db. The minimum rejection for a flanked filter is about 42 db. When several receiving channels are connected at one terminal, the difference in receiving levels between adjacent channels should not be greater than 15 db. This level difference will cause distortion of 2 to 3% at speeds of 100 words per minute. Smaller level differences should be used if at all possible, preferably not exceeding 5 db. Level differences of approximately 15 db may cause false operation of the supervisory feature on an adjacent channel when this option is used. For frequencies separated by two or more channel assignments, this level difference is somewhat less critical.

Because the per-channel receiver input level can be as low as −40 dbm, and the combined transmitting level can be as high as +6 dbm, the total line loss between transmitter and receiver can be as great as 46 db, line noise conditions permitting. The maximum permissible line loss will decrease as the number of channels in a system increases.

System Application Considerations

Operation on Carrier or Radio

Perhaps the most common application of telegraph carrier is on voice-frequency (v-f) carrier or radio facilities. This is usually on a 4-wire basis, since this permits the most efficient use of the available bandwidth. Also, these facilities are by their nature a 4-wire medium with

Vol.
Pt. 1
195

Table III. Single-Channel Transmitting Level in dbm for Different Numbers of Telegraph Channels for Various Facilities

Total No. of Telegraph Channels	Type of Facility, Maximum Per-Channel Levels			
	4-Wire Open Wire or Cable	2-Wire Open Wire or Cable	−13 dbm Cxr. Channel	−16 dbm Cxr. Channel
1	+6.0	+2.0	−16.0	−19.0
2	+3.0	−1.0	−22.0	−25.0
3	+1.2	−2.8	−25.5	−28.5
4	0	−4.0	−29.0	−32.0
5	−1.0	−5.0	−30.0	−33.0
6	−1.8	−5.8	−31.0	−34.0
7	−2.5	−6.5	−31.5	−34.5
8	−3.0	−7.0	−32.0	−35.0
9	−3.5	−7.5	−32.5	−35.5
10	−4.0	−8.0	−33.0	−36.0
11	−4.4	−8.4	−33.4	−36.4
12	−4.8	−8.8	−34.0	−37.0
13	−5.2	−9.2	−34.0	−37.0
14	−5.5	−9.5	−34.0	−37.0
15	−5.8	−9.8	−34.0	−37.0
16	−6.0	−10.0	−34.0	−37.0
17	−6.3	−10.3	−34.0	−37.0
18	−6.5	−10.5	−34.0	−37.0
19	−6.8	−10.8		
20	−7.0	−11.0		
21	−7.2	−11.2		
22	−7.4	−11.4		
23	−7.6	−11.6		
24	−7.8	−11.8		
25	−8.0	−12.0		
26	−8.1	−12.1		

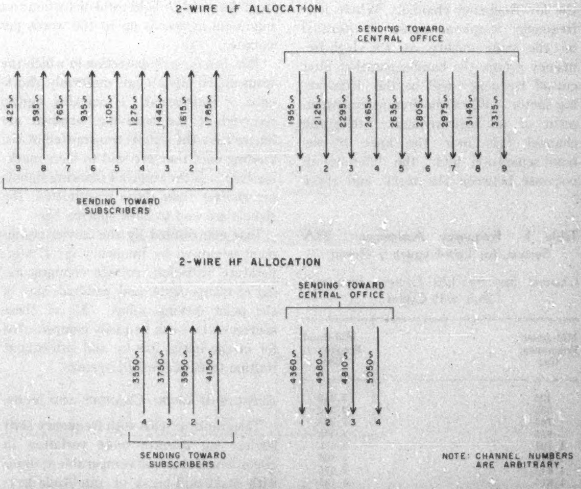


Fig. 5. Typical frequency allocations for channels operating on 2-wire lines

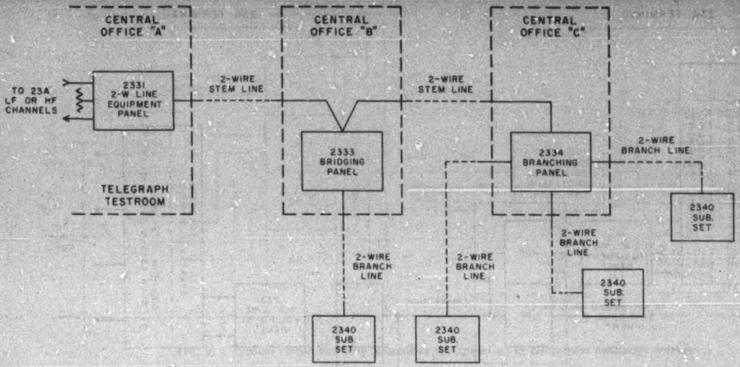

CENTRAL OFFICE "A"
TO 23A LF OR HF CHANNELS
2331 2-W LINE EQUIPMENT PANEL
TELEGRAPH TESTROOM
2-WIRE STEM LINE
CENTRAL OFFICE "B"
2333 BRIDGING PANEL
2-WIRE STEM LINE
CENTRAL OFFICE "C"
2334 BRANCHING PANEL
2-WIRE BRANCH LINE
2340 SUB. SET
2-WIRE BRANCH LINE
2340 SUB. SET
2-WIRE BRANCH LINE
2340 SUB. SET
2-WIRE BRANCH LINE
2340 SUB. SET

Fig. 6. Block diagram showing typical layout of 2-wire low-frequency or high - frequency channels

the transmitting and receiving paths fully separated. As described in the preceding section, the number of telegraph channels will depend on the bandwidth of the transmitting facility. The individual telegraph carrier channel transmitting level will be determined by the number of channels required in the telegraph system and the allowable voice-frequency channel input level of the particular carrier system used.

Assuming a v-f channel test tone input level of −16 dbm, the *23A* per-channel transmitting level would be set at −37.0 dbm on the basis of a full system of 18 telegraph channels. Table III shows permissible variation in maximum per-channel level with different numbers of channels in a system.

Note that the permissible per-channel level decreases more rapidly on v-f carrier channels than when using physical circuits. This is because the power-handling capacity of the v-f carrier channel is the limiting factor. This peak loading is not so critical on physical circuits. On carrier channels, the permissible per-channel level decreases on a logarithmic curve related to peak voltage, not average power, as in the case of physical circuits.

A fixed per-channel level generally is adopted as standard when several telegraph carrier systems terminate at one point. A per-channel tone level 21 db below the test tone level of the v-f carrier channel could probably be safely used for any number of channels up to the capacity of the v-f channel. Where only a few telegraph systems are operating on channels of a v-f carrier, this should not overload the v-f carrier system.

When the ultimate number of telegraph channels in a system will be less than 12, it may be desirable to operate the telegraph channels at as high a per-channel level as possible for the best signal-to-noise ratio.

Level Co-ordination with Low-Frequency Telephone Carrier Channels

There are two types of situations which require co-ordination between levels of telegraph systems and levels of low-frequency telephone carrier systems. Operating levels for the telegraph channels must be considered with relation to low-frequency telephone carrier channels operated on: (1) the same pair and (2) adjacent pairs.

Consider the case where *23A* channels operate on a physical circuit at frequencies occupied on adjacent pairs by low-frequency telephone carrier channels. Because the telegraph channels could crosstalk into the adjacent pairs, the transmitting level of the telegraph channels must be held below the point at which objectionable crosstalk could occur.

The most direct method of determining the allowable transmitting level for the *23A* system is to measure the near-end and far-end crosstalk levels between the circuit proposed for the *23A* and adjacent circuits which could be disturbed. For the low-frequency channels, crosstalk should be measured at 1,000 cps. A crosstalk coupling loss of 55 db may generally be taken as acceptable.

For *23A* systems operating in open-wire lines, sending levels will be governed by the type of transpositions in the line, the frequencies of carrier circuits, and the positions of these circuits with respect to the pin positions of the *23A* circuits. The effect of these conditions can be determined by measuring crosstalk coupling values on disturbed circuits. For the high-frequency channels, crosstalk measurements should be made at a frequency of approximately 4,500 cycles.

Where carrier circuits are operated in the same open-wire lead, the *23A* levels should be co-ordinated, insofar as possible, with other carrier levels to minimize the

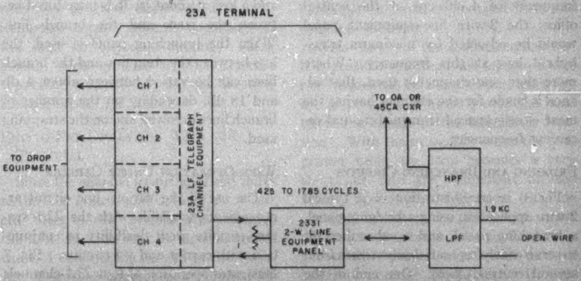

23A TERMINAL
CH 1
CH 2
CH 3
CH 4
TO DROP EQUIPMENT
23A LF TELEGRAPH CHANNEL EQUIPMENT
425 TO 1785 CYCLES
2331 2-W LINE EQUIPMENT PANEL
TO OA OR 45CA CXR
HPF
1.9 KC
LPF
OPEN WIRE

Fig. 7. Block diagram of typical terminal with four low-frequency channels on open-wire operating on 2-wire basis below Lenkurt 45CA or Western Electric OA carrier

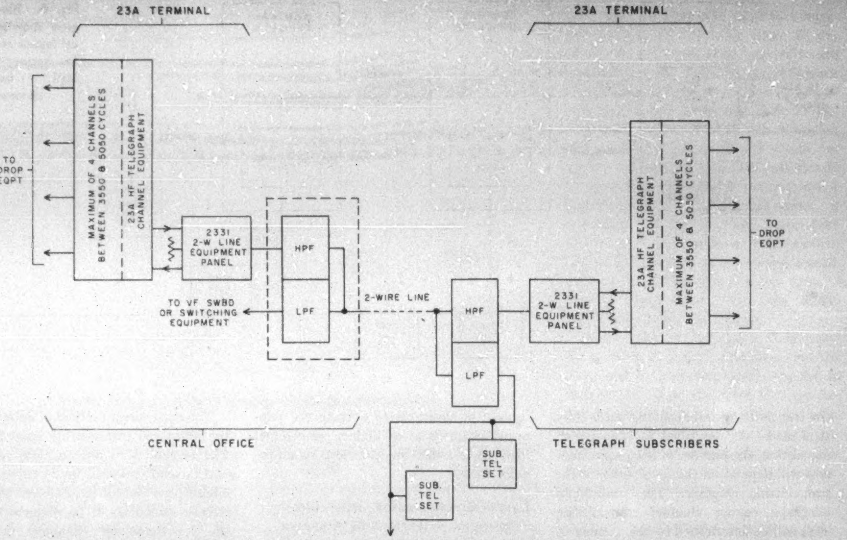

23A TERMINAL
23A TERMINAL
TO DROP EQPT
MAXIMUM OF 4 CHANNELS BETWEEN 3550 & 5050 CYCLES
23A HF TELEGRAPH CHANNEL EQUIPMENT
2331 2-W LINE EQUIPMENT PANEL
HPF
LPF
TO VF SWBD OR SWITCHING EQUIPMENT
2-WIRE LINE
HPF
LPF
2331 2-W LINE EQUIPMENT PANEL
23A HF TELEGRAPH CHANNEL EQUIPMENT
MAXIMUM OF 4 CHANNELS BETWEEN 3550 & 5050 CYCLES
TO DROP EQPT
CENTRAL OFFICE
TELEGRAPH SUBSCRIBERS
SUB. TEL SET
SUB. TEL SET

Fig. 8. Block diagram showing typical layout of 2-wire high-frequency channels operating above the voice band on subscriber telephone lines

Vol.
Pt. 1
1959

crosstalk into any physical or carrier v-f circuit operating in the lead.

Two-Wire Transmission

Fig. 5 shows a typical frequency assignment for 2-wire operation. In order to transmit and receive on the same 2-wire circuit, it is necessary to combine the transmit and receive branches of the telegraph carrier circuit where they connect to the 2-wire line circuit. This is accomplished by means of a transformer with a balanced winding. The center tap of the balanced winding is connected through a balancing network.

Assuming the use of channel *1* of the high-frequency allocation with a sending frequency of 4,150 cps at the central office, the 2-wire line-equipment panel would be adjusted for maximum trans-hybrid loss at this frequency. Where more than one channel is used, the balance is made for the channel having the most closely spaced transmitting and receiving frequencies.

Bridging and Branching Circuits

Fig. 6 is an illustration of a typical 2-wire application using a bridging panel, a branching panel, and a subscriber set to serve widely separated subscribers from several central offices. One end of the telegraph carrier circuit terminates in the telegraph test room.

Fig. 6 also illustrates use of the terms "stem line" and "branch line." The stem line generally starts at the central office and applies to the section of the system carrying more than one channel away from the central office. The branches feed from the stem and distribute the circuits to the various subscribers' locations. A branch may carry more than one channel, and there may be instances where it is difficult to define just where the stem ends and a branch begins, since all channels but one may have left the stem as it continues to the farthest subscriber.

When the bridging panel is used, a loss of 1 db is introduced into the stem line, and a loss of about 16 db is introduced between the stem and the branch line. When the branching panel is used, the loss between the stem line and the branch lines can be varied between about 4 db and 18 db, depending on the number of branch lines involved and on the strapping used.

Wire Operation Under Carrier

The use of the various line circuit arrangements available with the *23A* system permits great flexibility in conjunction with carrier and v-f circuits. Fig. 7 illustrates operation of four *23A* channels on a 2-wire basis underneath the Lenkurt *45CA* or Western Electric *OA* system using separation filters. A band separation filter with the crossover at 1,900 cps divides the 2-wire line for the v-f telephone and *23A* telegraph carrier frequencies. Because of the low filter crossover point required by these telephone carrier systems, only frequencies between 425 and 1,785 cycles may be used for telegraph.

When the *23A* system is used in this application, the direction of transmission is co-ordinated with the *OA* or *45CA* system so that the low group transmitting (LGT) channels of the *OA* or *45CA* send in the same direction as the higher frequency channels (1,275 to 1,785 cycles) of the *23A* system. This places adjacent frequencies at more nearly the same levels and helps to minimize crosstalk.

Two-Wire Telegraph Superimposed on Subscriber Telephone

Fig. 8 illustrates a method of applying the higher frequency channels of the *23A* system to subscriber manual or dial lines for carrying up to four channels of telegraph to an outlying point on a 2-wire basis, in addition to the usual number of subscribers' telephones on the same circuit. Cable attenuation will be the limiting factor at the higher frequencies. Open-wire circuits would permit operation over longer distances since less attenuation is offered to the higher frequencies. This type of operation is also

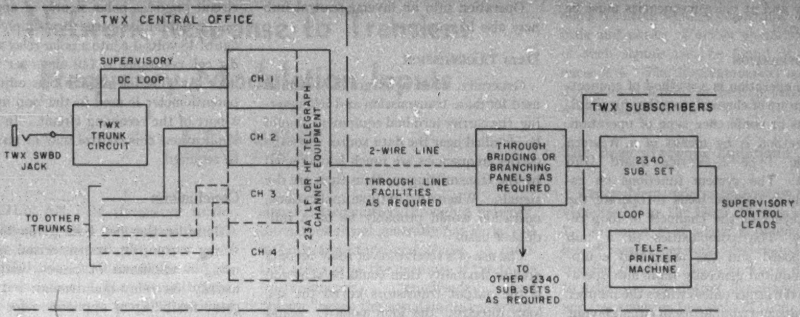

TWX CENTRAL OFFICE
SUPERVISORY
DC LOOP
TWX TRUNK CIRCUIT
TWX SWBD JACK
TO OTHER TRUNKS
CH 1
CH 2
CH 3
CH 4
23A LF OR HF TELEGRAPH CHANNEL EQUIPMENT
2-WIRE LINE
THROUGH LINE FACILITIES AS REQUIRED
THROUGH BRIDGING OR BRANCHING PANELS AS REQUIRED
TO OTHER 2340 SUB SETS AS REQUIRED
TWX SUBSCRIBERS
2340 SUB. SET
LOOP
TELE-PRINTER MACHINE
SUPERVISORY CONTROL LEADS

Fig. 9. Typical circuit showing subscriber set operating in teleprinter exchange service

called "speech-plus." Proper operation in applications of this type will also depend on other factors, one of which is potential interference between the telegraph and other services on the same line or cable.

This method of superimposed carrier operation provides an economical use of circuits, particularly where physical facilities may not be too plentiful.

In this "speech-plus" application, either a filter having a 1.9-kc separation frequency or a 3.0-kc separation frequency may be used. In either case, the lowest telegraph channel frequency will be governed by the filter crossover frequency, with due consideration of the 1 db mark/space level difference for telegraph covered in the preceding section.

Teleprinter Exchange Service With 2-Wire Carrier Operation

Fig. 9 illustrates teleprinter exchange service in 2-wire operation using the subscriber set. When the *23A* system is used in 2-wire service, it is advisable to assign the channel frequencies so that the high-frequency terminals send toward the central office. As telephone circuits get farther away from the central office, they are more exposed to noise-producing sources, such as power lines and lightning. The circuits near the central office are shielded by other circuits in the same cable, and by the cable sheath itself. Also, much of the route may be underground.

The higher channel frequencies are attenuated more than the lower channel frequencies. For a given circuit and channel frequency assignment, one direction of transmission will have more attenuation than the other. Therefore, it is advisable to assign the frequency having the greatest attenuation in the direction which is least subject to noise at the receiving terminal.

Bridging panels and/or branching panels are located at intermediate points as required to serve branch lines to intermediate subscribers.

Where line characteristics will permit, all the frequencies from 425 to 5,050 cps may be used. Typical assignments are shown in Fig. 5. These may, of course, be varied to suit local requirements; the channel filter networks would be selected to agree with the required frequencies.

Frequency and Direction Considerations

When laying out a telegraph carrier system which is to operate on the same facilities with open-wire carrier systems, the telegraph carrier frequencies and directions of transmission should be coordinated with the telephone carrier frequencies. Adjoining bands should always transmit in the same direction to minimize crosstalk.

Determination of the transmission level for the telegraph system will involve crosstalk and frequency considerations with any program circuits and/or low-frequency carrier systems operating on the same or adjacent circuits. These include such systems as the Lenkurt types *32E*, *33A*, *45CA*, and the Western Electric types *C*, *H*, and *OA*.

End-to-End Operation With 43A Terminal

Since the channel frequencies and shifts of the *23A* system and those of the Western Electric *43A* system are identical, the *23A* channels may be operated with the *43A* channels on an end-to-end basis. Any frequency combination required by the *43A* system may be met with the *23A* system. Channel operating levels and supervisory on-off functions are also compatible.

Terminating and D-C Loop Arrangements

Direct-Current Loops

Operation of telegraph carrier channels is closely related to operation of d-c telegraph circuits. Selection of the proper d-c loop termination is just as important to end-to-end circuit operation as is the proper application of telegraph carrier channels to various types of physical or carrier-derived circuits. The type of loop facility available, variable loop line conditions caused by weather, and the type of teleprinter service required are all determining factors in choosing the best loop option for a particular condition.

The loop may be two conductors, or may consist of one wire with ground as a return circuit. Full flexibility is offered; neutral, polar, half duplex, full duplex, inverse neutral, and any combination of battery voltages and polarities may be employed as required. By the simple process of changing a strapping plug, the loop operation can be changed from one option to the other.

With the use of the *2340* subscriber set, it is possible to extend the telegraph carrier frequencies on a speech-plus basis (by use of proper line filters) over the local telephone circuit to the teleprinter location. With the subscriber set at the teleprinter table, the loop becomes a very short one into the equipment on the table. Speech-plus operation can be obtained by use of line filters with a cutoff frequency from 1.9 kc to 3.5 kc together with proper selection of data or telegraph carrier frequencies. If considerable cable is involved the difference in the loss at the telegraph carrier frequency and at voice frequencies must be considered.

HUB OPERATION

Hub operation is a method of connecting a group of channels together for simultaneous or conference type of operation. One method is by means of a Western Electric Type *2* Service Board Hub Circuit. The system functions on essentially a voltage basis to operate the send keyers of the channels or loop repeaters. The complexities of a hub switchboard and the associated equipment required generally limit hub operation to the larger offices where the number of circuits and complexity of circuit layout make this type of operation economical.

When Lenkurt *23A* channels work into a hub board, auxiliary coupling units are not required to couple the send and receive hub circuitry into the channels. The send and receive legs are connected directly to the corresponding terminals on the hub equipment panel when operating on either a half- or full-duplex basis.

Operation into an inverse-neutral hub may also be employed.

DATA TRANSMISSION

Generally, when a telegraph system is used for data transmission and telemetering, the carrier terminal equipment should be installed near the data source since the carrier frequencies are much less affected by the transmitting medium than are d-c signals. When used in these applications, operation would probably be on a full-duplex basis.

The use of a receive keyer relay permits greater flexibility than could be achieved if the output transistors keyed the d-c loop directly; the loop polarity would necessarily remain fixed for a given transistor configuration.

In specialized data applications, where it may not be necessary to use the relay of the loop equipment unit, signals may be taken from the d-c keyer transistors in the channel unit. This will require furnishing 48 volts of battery through the signal receiving device. The normal operating condition of the transistor keyer circuit provides polar signals of approximately 20 milliamperes at a supply voltage of 48 volts d-c into a polar relay whose d-c coil resistance is 140 ohms per winding. An external receive bias adjusting potentiometer is used in the loop unit as a part of the receiving circuit. In other applications this control may or may not be required.

Conclusions

Consideration has been given to providing a compact, transistorized system using a minimum of power, with unusually low relay maintenance but with great flexibility of application for data, telemetering, or telegraph transmissions. With the growing demand for data and telegraph circuits, Lenkurt's type *23A* Datatel System offers a logical solution to the problem of meeting these needs.

Reference

1. DESIGN FEATURES OF A NEW FREQUENCY SHIFT CARRIER TELEGRAPH SYSTEM, M. L. Stephens. *AIEE Transactions*, see pp. 927–34 of this issue.

◆

Discussion

A. B. Covey (American Telephone and Telegraph Company, New York, N. Y.): With reference to the introduction, "change in waveshape" is not necessarily distortion. Transmission of rectangular pulses always involves some rounding of the pulse, since there is some band limiting involved. Distortion is defined (American Standard *C*42.65) as time displacement of signal transitions with respect to their proper relative positions in perfectly timed signals. Since the signals are received by an electric circuit, rounded pulses are considered to be undistorted if the point on the transition at which the receiver responds occurs at the proper time.

Under the heading "Effects of Level Changes and Noise," it is stated that "input level changes of 35 db will produce less than 2% change in receiving bias." To be strictly correct, the stipulation should be added that this effect applies only if the level to which the signal drops is not so low as to be below the point at which limiting occurs.

The discussion under "Signal Level Limits" tends to be misleading. It is true that the levels employed on telephone carrier channels are governed by the total power-carrying capacity of the channel. However, where tones are transmitted on physical wire facilities it is noise influence considerations which limit the levels used. Noise influence is a function of the influence of the disturbing circuit, the susceptiveness of the disturbed circuit, and the loss in the coupling path. Each of these categories has variables within it, but since only the first is being discussed, it is a function of: (1) the level of the disturbing tone; (2) the frequency composition of the disturbing tone; and (3) the balance to ground of the disturbing circuit over the frequency range. Where multiple tones are used on the disturbing circuit, the interference depends on the total power of the combined tones after they have been weighted to account for the varying interference value of different frequencies of speech currents or program currents, as the case may be.

With regard to the actual levels tabulated in Table III, for one thing, the first two columns do not specify at what transmission level point the figures apply. For another, they do not specify whether the tones are being applied to an exchange cable or a toll cable. If we assume it to be the zero level point, which is the common method of specifying such levels, and that both columns refer to a toll grade facility, then the levels are considerably higher than those which the Bell System uses. For example, for the 17-channel case we use a level of −13 dbm per channel (at the 0 level point) if there are no program circuits in the exposed group and −23 dbm if program circuits are included. These levels are derived from the noise influence considerations described previously.

For less than 17 channels, any increase above these levels depends on which of the 17 frequencies are being used. For example, for a single channel we would allow a level of +13 dbm for the 425-cps channel, but only −16 dbm for the 3,145-cps channel from the standpoint of interference into nearby voice circuits.

With regard to the levels given in Table III for telephone carrier channels, it should be pointed out that the average power of a steady tone is about 8 db higher than the average power of a voice signal. For equivalent loading of the carrier system, then, the total power of the combined telegraph tones should be kept about 8 db below the normal telephone level. This means that for 16 channels, the single-channel level would have to be −36 dbm at the −16-dbm level point. For one channel, the corresponding level would be −24 dbm at the same point.

There is only one other comment, and that is on Fig. 8. It is assumed that this application to subscriber telephone lines can only be used for the case of a "dry" line, that is, a line using local battery subscriber's telephone set. It will not apply to the normal common battery subscriber loop carrying direct current.

T. B. Collins, Jr., and J. E. Pitts: The comments of Mr. Covey are appreciated. As indicated in his discussion and in the paper, noise and crosstalk considerations are of major concern in the operation of any carrier system. The particular levels chosen must take into account the characteristics of the facility over which the signals are to be transmitted. The need for standardization of operating practices in large companies sometimes makes it advisable to specify the lower operating levels to assure satisfactory performance with a minimum of engineering effort on each project. Provided that crosstalk performances for the particular facility are within proper limits, the higher levels will permit better signal-to-noise margin.

Operation of the *23A* equipment on common battery subscriber lines is satisfactory when the proper Lenkurt line filters are used. In such applications it is essential that the line filters used be so designed that the direct current in the subscriber loop does not saturate the inductors in the filter. The Lenkurt line filters used in this application provide entirely satisfactory performance.

Network Response to Transient Frequency Modulation Inputs

T. T. N. BUCHER
ASSOCIATE MEMBER AIEE

WHEN AN F-M (frequency modulated) signal is transmitted through a linear passive network, the output signal is hybrid, having both amplitude and frequency variations. The output expression can be written in terms of the input through use of circuit equations. However, the relationship is sufficiently complex that an evaluation of a direct solution of these equations is not possible except in simple cases. When the input modulation is periodic, or has a discontinuous step or impulse form, direct solution of the equation is useful; however, in other cases, approximations prove helpful in obtaining a solution.

Approximations have been given for cases where the transfer function of the network differs only slightly from the ideal and may be approximated simply. When the modulation is a relatively slowly varying signal, the "quasi-stationary" approximation and its many variations are useful.[1] However, when the network differs appreciably from the ideal and when the modulation varies rapidly, these techniques are of limited value. The present paper is concerned primarily with approximations which are useful for rather general network functions, when the input modulation varies rapidly during a relatively short period. Such modulations are encountered, for example, when television signals or pulse modulation signals are used to frequency modulate a carrier.

Instantaneous Amplitude and Phase

For purposes of analysis, it is convenient to represent the actual input and output signals of a network as the real parts of phasor functions. When an F-M signal is transmitted through a network, the output signal generally has variations in both amplitude and phase. A question arises as to the meaning of the terms (instantaneous) "amplitude" and "phase," when the signal under consideration is not of simple sinusoidal form.

A real function of time can be represented as the real part of any of an infinity of phasor functions formed by adding arbitrary imaginary functions to the original real function. Unless some restriction is placed on the imaginary function, the terms, amplitude and phase are relatively meaningless. For the purposes of this paper, amplitude and phase are defined as the magnitude and angle of that phasor function associated with the real function which has been called the "analytic" signal.[2-4] This signal is formed by adding to the real signal $f(t)$, a quadrature signal $g(t)$, derived from it by the Hilbert transform relationship[5]

$$g(t)=-\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{f(\tau)}{\tau-t}d\tau \qquad (1)$$

The resulting phasor function

$$\psi(t)=f(t)+jg(t)=a(t)\epsilon^{j\theta(t)} \qquad (2)$$

is analytic throughout the upper half of the complex time plane. On the real time axis, its real part is the original signal $f(t)$. The spectrum (fourier transform) of $\psi(t)$ is zero on the negative real frequency axis and on the positive real frequency axis, has twice the value of the spectrum of the original real time function $f(t)$. Conversely, signals whose spectra are zero on the negative real frequency axis, are analytic signals.

When an analytic signal is applied to the input of a linear passive network, the output signal is analytic. The amplitude and phase of the analytic signal reduce to the usual definitions, when $f(t)$ is a sinusoidal function. When the real signal is a narrow band signal (spectrum spread small compared to center frequency), the amplitude and phase functions satisfy one's intuitive feelings concerning amplitude and phase. Further characteristics of such signals can be found in references 2-4. The (instantaneous) angular frequency is defined as the time derivative of the phase.

The Basic Network Equations

The problem under consideration can be visualized from the block diagram in Fig. 1. An input F-M signal is applied to a linear passive network (designated by a transfer function $T(\omega)$), and it is desired to determine the output signal amplitude and frequency as functions of time.

When the input signal is $\epsilon^{j\omega_0 t}\epsilon^{j\Delta\omega_p H(t)}$ a normalized output signal may be obtained by use of the superposition theorem as

$$\beta(t)\epsilon^{j\theta(t)}=\int_0^{\infty}\int_{-\infty}^{\infty}\epsilon^{j\Delta\omega_p H(\tau)}\epsilon^{j(\omega-\omega_0)(t-\tau)}\times \frac{T(\omega)}{T(\omega_0)}d\tau d\omega \qquad (3)$$

$$=\int_{-\infty}^{\infty}\epsilon^{j\Delta\omega_p H(\tau)}u(t-\tau)d\tau \qquad (4)$$

where

$\beta(t)$ = output amplitude divided by $|T(\omega_0)|$
$\theta(t)$ = output phase deviation from the reference phase, $\omega_0 t+\underline{/T(\omega_0)}$
$H(t)=\int_{-\infty}^{t}h(t)dt$
$\Delta\omega_p h(t)$ = angular frequency deviation from the reference angular frequency, ω_0
$\Delta\omega_p$ = peak angular frequency deviation
$u(t)=\int_0^{\infty}\epsilon^{j(\omega-\omega_0)t}\frac{T(\omega)}{T(\omega_0)}d\omega=u_r(t)+ju_i(t)$

The integration for $u(t)$ has the lower limit zero because of the assumption that input signal is analytic. When the inner integration is performed in equation 3, the resultant function is zero for negative frequencies. Hence only the integration along the positive frequency axis is required. Thus, $u(t)$ may be considered a normalized analytic impulse response of the network. (The equations can be used if the signal is not analytic, by extending the lower limit of the first integration in equation 3 to minus infinity, and allowing $u(t)$ to be the usual impulse response. However, the ambiguity in meaning of "amplitude" and "phase" then exists.)

The angular frequencies and time variable can be normalized by dividing the

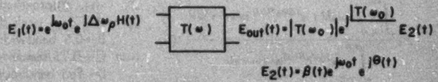

E1(t)=e^{jω0t}e^{jΔωpH(t)}
T(ω)
Eout(t)=|T(ω0)|e^{j/T(ω0)} E2(t)
E2(t)=β(t)e^{jω0t}e^{jθ(t)}

Fig. 1. Block diagram of network defining input and output functions

---

Paper 59-1144, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11-16, 1959. Manuscript submitted April 17, 1959; made available for printing August 24, 1959.

T. T. N. BUCHER is with the Radio Corporation of America, Moorestown, N. J.

This paper is based on portions of a dissertation submitted by the author in partial fulfillment of the requirements for the Ph.D. degree at the University of Pennsylvania. The author wishes to acknowledge the assistance and suggestions offered by J. G. Brainerd, R. S. Berkowitz, and R. M. Showers, in the development of the work.

Table I. Coefficients in Expansion for Amplitude and Frequency

m	$B_m(t)$	$A_m(t)$
0	1	$K_1'(t)$
1	$K_1(t)-K_1^*(t)$	$K_2'(t)/2!-K_1'(t)K_1(t)$
2	$K_2(t)-2K_1(t)K_1^*(t)+K_2^*(t)$	$K_3'(t)/3!-K_2'(t)K_1(t)/2!-K_1'(t)K_2(t)/2!+K_1'(t)K_1^2(t)$
3	$K_3(t)-3K_2(t)K_1^*(t)+3K_1(t)K_2^*(t)-K_3^*(t)$	$K_4'(t)/4!-K_3'(t)K_1(t)/3!-K_1'(t)K_3(t)/3!-K_2'(t)K_2(t)/(2!)^2+K_2'(t)K_1^2(t)/2!+$ $K_1'(t)K_1(t)K_2(t)-K_1'(t)K_1^3(t)$
4	$K_4(t)-4K_3(t)K_1^*(t)+6K_2(t)K_2^*(t)-4K_1(t)K_3^*(t)+K_4^*(t)$	

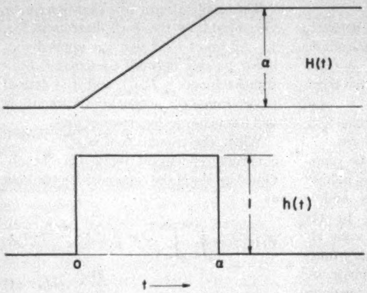


Fig. 2. Rectangular frequency modulating function, h(t), of duration α, and resulting phase modulating function H(t)

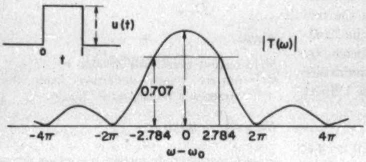


Fig. 3. Magnitude of transfer function $|T(\omega)|$, of network having rectangular impulse response of unity amplitude and unity duration

former and multiplying the latter by an arbitrary factor with dimensions of angular frequency. In the examples presented later the normalizing factor is half of the 3-db (decibels) bandwidth (expressed in angular frequency) of the circuit under study.

By multiplying equation 4 by its complex conjugate, an expression for the square of the amplitude can be obtained

$$\beta^2(t)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\epsilon^{j\Delta\omega_p[H(\tau)-H(p)]}\times u(t-\tau)u^*(t-p)d\tau dp \quad (5)$$

u^* = complex conjugate of u

Since, generally, when a complex function z, has a derivative with respect to t, $d(\text{Arg } z)/dt = Im\ dz/zdt$, normalized angular frequency of the output can be expressed as

$$\frac{d\theta(t)}{dt}=Re\frac{\int_{-\infty}^{\infty}\Delta\omega_p h(\tau)\epsilon^{j\Delta\omega_p H(\tau)}u(t-\tau)d\tau}{\int_{-\infty}^{\infty}\epsilon^{j\Delta\omega_p H(\tau)}u(t-\tau)d\tau} \quad (6)$$

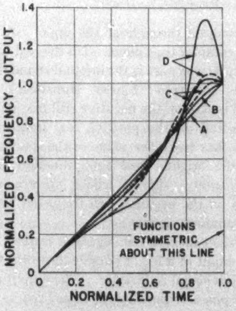


Fig. 4. Output frequency response of network with rectangular impulse response to F-M signal with rectangular modulation of same width as impulse response

—— Exact solution
- - - - Approximation
A—Zero peak deviation
B—0.5 bandwidth peak deviation
C—0.75 bandwidth peak deviation
D—Full bandwidth peak deviation

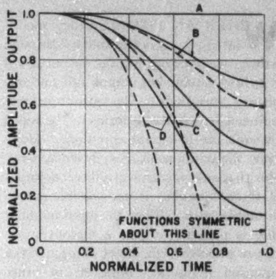


Fig. 5. Output amplitude response for same conditions as in Fig. 4

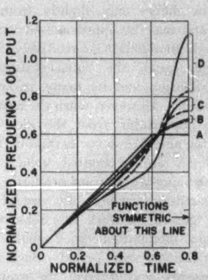


Fig. 6. Similar to Fig. 4 except that input modulation has 0.6 width of impulse response

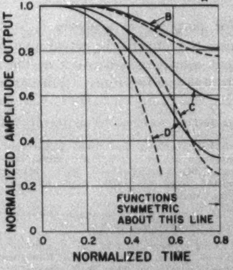


Fig. 7. Output amplitude response for same conditions as in Fig. 6

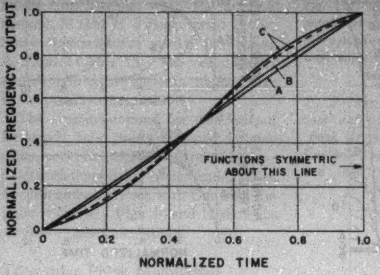


Fig. 8. Output frequency response of network with rectangular impulse response to F-M signal with modulation of same width as impulse response and zero peak deviation

A—No detuning
B—±0.25 bandwidth detuning
C—±0.5 bandwidth detuning

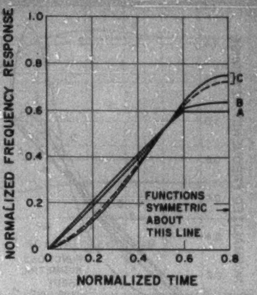


Fig. 9. Similar to Fig. 8 except that input modulation has 0.6 width of impulse response of network

Evaluation of the Output Functions

The solution of the problem resides in the evaluation of the equations 5 and 6. In the past, a number of approaches have been used in performing this evaluation. The integrals can be evaluated directly in some simple cases: however, in many cases, some form of approximation has been employed.

If $H(t)$ is periodic, the first factor under the integrals in equation 3 can be expanded in an infinite trigonometric series. The response to each term of the series can be determined from the transfer function $T(\omega)$, and the resultant terms may be summed to provide as high a degree of accuracy as desired. In general, the output series will not be converted easily to closed form; however, in some cases, such a technique has proved effective.[8]

If, in equation 3, the form of $T(\omega)$ can be approximated simply, the equation may be evaluated without undue difficulty. In some cases, the magnitude of the transfer function has been assumed to vary from unity by only a small amount and the phase from linearity by a small amount.[7] The output then becomes equal to a delayed input plus perturbations resulting from the small variations in the transfer function.

A well-known approximation due to Carson and Fry[1] is obtained by expanding the exponential term in equation 4 as a power series in τ (and performing additional manipulation). This is the so-called quasi-stationary solution, which is useful when the modulation varies relatively slowly. While it can be shown that the resultant series is not generally convergent,[8] it may be used with a finite number of terms. Usually only the first two terms are employed.

Expansion for Rapidly Varying Modulation

When the modulating function varies rapidly, as in a pulse-like or step-like function, but the total phase has small variation, a useful expansion is obtained by expanding the equations 4–6 as power series in $\Delta\omega_p$. (The approximations are useful for any case where the total phase variation is small.) If $u(t)$, $h(t)$, and $H(t)$ are bounded, and differentiation under the integral sign is permitted, the resultant series converge. In this case, the various output functions can be expressed as follows (since developing these expressions, the author has become familiar with reference 9 which gives the same results for equation 9)

$$\beta(t)\epsilon^{j\theta(t)} = \sum_{m=0}^{\infty} \frac{(j\Delta\omega_p)^m}{m!} K_m(t) \tag{7}$$

$$\beta^2(t) = \sum_{m=0}^{\infty} \frac{(j\Delta\omega_p)^m}{m!} B_m(t) \tag{8}$$

$$\frac{d\theta(t)}{dt} = \Delta\omega_p Re\left[\sum_{m=0}^{\infty} (j\Delta\omega_p)^m A_m(t)\right] \tag{9}$$

where

$$\left.\begin{aligned} K_m(t) &= \int_{-\infty}^{\infty} H^m(\tau)u(t-\tau)d\tau \\ K_0(t) &= \int_{-\infty}^{\infty} u(t-\tau)d\tau = 1 \end{aligned}\right\} \tag{10}$$

$$B_m(t) = \sum_{r=0}^{m} (-1)^r \binom{m}{r} K_{m-r}(t)K_r^*(t) \tag{11}$$

$$A_m(t) = \begin{bmatrix} 1 & 0 & 0 & \ldots K_1'(t) \\ K_1(t) & 1 & 0 & \ldots \frac{K_2'(t)}{2!} \\ \frac{K_2(t)}{2!} & K_1(t) & 1 & \ldots \frac{K_3'(t)}{3!} \\ \frac{K_3(t)}{3!} & \frac{K_2(t)}{2!} & K_1(t) & \ldots \frac{K_4'(t)}{4!} \\ \vdots & & & \\ \frac{K_m(t)}{m!} & \frac{K_{m-1}(t)}{(m-1)!} & \frac{K_{m-2}(t)}{(m-2)!} & \ldots \frac{K_{m+1}'(t)}{(m-1)!} \end{bmatrix} \tag{12}$$

The first several coefficients, $B_m(t)$ and $A_m(t)$ are shown in Table I.

Remainder

The expansions can be terminated after a finite number of terms. Using Taylor's formula with the remainder, expressions can be determined from which the bounds of the resulting error can be evaluated. The expressions are so complex, however, that the bounds calculated are not very satisfactory. Consequently, some cases have been chosen where the original equations could be solved in closed form, and the results have been compared with the expansions using terms only up to the second order in $\Delta\omega_p$ in the approximations.

Rectangular Pulse or Step Modulation

If the modulation is a rectangular pulse in frequency, as indicated in Fig. 2, and if the network considered has a rectangular impulse response (transfer function is indicated in Fig. 3), it is possible to integrate the complete equations 5 and 6. A number of calculations have been made, for various widths of pulse, various modulation amplitudes and various degrees of detuning from the center frequency of the

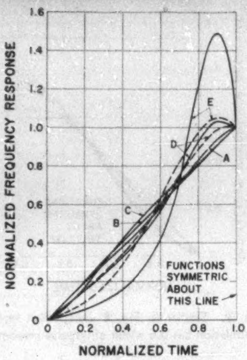


Fig. 10. Similar to Fig. 8 except that peak deviation is one-half bandwidth

A—No detuning
B—Detuning of −0.25 bandwidth
C—Detuning of −0.5 bandwidth
D—Detuning of +0.25 bandwidth
E—Detuning of +0.5 bandwidth

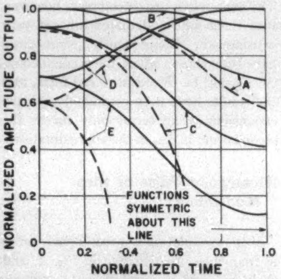


Fig. 11. Amplitude output response for same conditions as Fig. 10

filter. In each case, the output frequency and amplitude have been calculated, employing equations 5 and 6, and also employing up to second order terms of equations 8 and 9. (Cases employing this same input have been evaluated in reference 10 which came to the author's attention after completing this work. Ditl has developed a number of additional cases.) Where detuning is encountered, approximations are obtained with more terms than occur in equations 8 and 9.

Some typical results are shown in Figs. 4–13. For deviation approaching zero, the output frequency is identical with the amplitude output for a similarly amplitude-modulated signal applied to the same circuit.

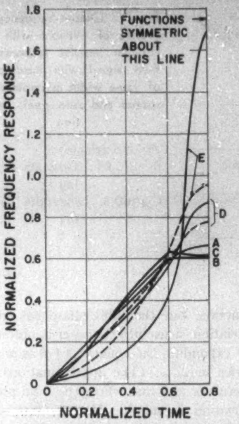


Fig. 12. Similar to Fig. 10, except that modulation has 0.6 width of impulse respnose

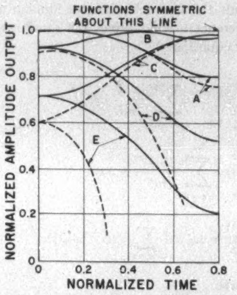


Fig. 13. Output amplitude response for same conditions as Fig. 12

The results suggest that for inputs of this general nature, the approximation for output frequency, using no higher than second-order terms, is good within a few percent so long as the peak deviation plus detuning stays within the 3-db bandwidth of the circuit. If 10% error is tolerable, the sum of these terms may approach $1^{1}/_{2}$ times this bandwidth.

For the amplitude approximation, using no higher than the second-order term, greater error is encountered. Within the 3-db bandwidth, errors of the order of 10% are encountered; for only a few percent error, the sum of detuning and peak deviation should remain within about 3/4 of the 3-db bandwidth.

Since these results have been established for only one class of pulses and circuit, the foregoing indications can only be considered to be guides, rather than firm estimates. These indications are about 6 times the allowable sum of detuning plus deviation which has been calculated in a few cases using the upper bounds for the remainder calculated from Taylor's formula.

In addition to the foregoing, similar calculations have been made for the same

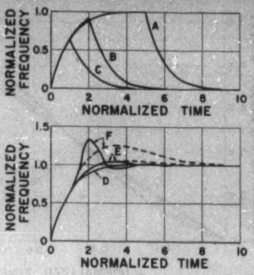


Fig. 14. Output frequency response of single-tuned circuit to F-M signal having rectangular modulation

A—Input modulation of 5 normalized time units width and one-fourth bandwidth peak deviation
B—Two normalized time units width and one-half bandwidth peak deviation
C—One normalized time unit width and full bandwidth peak deviation
D, E, F—Step function inputs, and one-quarter, one-half, and full bandwidth peak deviation respectively
- - - - Approximation differs significantly from complete solution

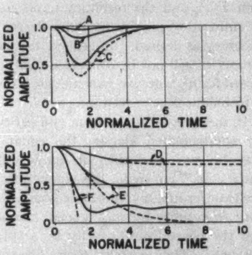


Fig. 15. Output amplitude response for same circuit and modulation type as Fig. 14

A, B, C—Unit width of modulation and peak deviations of one-quarter, one-half, and full bandwidth respectively
D, E, F—Step function input and same peak deviations respectively

Vo
Pt.
19

modulating function when the network response is that of a single-tuned circuit. Some representative results are shown in Figs. 14 and 15. The second-order approximations are better in this case than in the previous one.

The approximations for the output frequency have been calculated for two other network response characteristics—a network comprising two cascaded tuned circuits and a network with Gaussian impulse response. Figs. 16 and 17 show the results of these calculations.

Triangular Pulse Modulation

To determine the effects of input modulation without discontinuities, the response of the network with rectangular response to triangular input pulses has also been calculated using equations 8 and 9. In this case, while up to second-order terms have been used in making the frequency approximation, up to fourth-order terms have been used in making the amplitude approximation. The results are shown in Figs. 18–20. One is struck by the relatively small effect, on the output frequency waveform caused by varying peak deviation up to the 3-db bandwidth in this case.

Conclusions

An approximation technique has been presented for calculating response of a linear passive network to pulse-like and step-like frequency modulating functions. Results obtained by this technique indicate that so long as the frequency deviation remains within the 3-db bandwidth of the network, reasonably good approximations of output amplitude and frequency may be obtained using no higher than second order terms in the approximation series. In cases close to those often encountered in practice, there is surprisingly little difference in output frequency waveform, when the peak deviation varies from very small values up to full 3-db bandwidth. While the approximation formulas are applicable rather generally, care should be taken in extending these rough rules of approximation to cases differing appreciably from those presented.

References

1. VARIABLE FREQUENCY ELECTRIC CIRCUIT THEORY WITH APPLICATION TO THE THEORY OF FREQUENCY MODULATION, J. R. Carson, T. C. Fry. *Bell System Technical Journal*, New York, N. Y., vol. 16, 1937, p. 513 et seq.

2. THEORY OF COMMUNICATION, D. Gabor. *Journal*, Institution of Electrical Engineers, London, England, vol. 93, pt. III, 1946, p. 429 et seq.

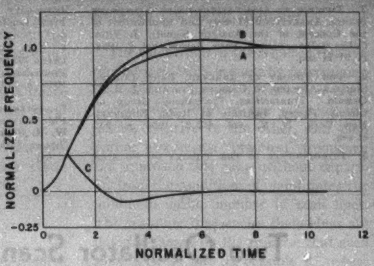


Fig. 16. Approximations to output frequency response of two cascaded tuned circuits to F-M signal having rectangular modulation

A, B—Step modulation with zero and one-half bandwidth peak deviation respectively
C—Unit width modulation and half bandwidth peak deviation

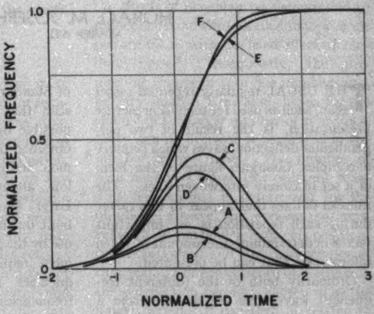


Fig. 17. Approximation to output frequency response of circuit having Gaussian impulse response to F-M signal having rectangular modulation

A, B—Input modulation of 0.3 normalized time units width and zero and half bandwidth peak deviation respectively
C, D—0.831 units width, zero, and half bandwidth respectively
E, F—Step input, zero, and half bandwidth respectively

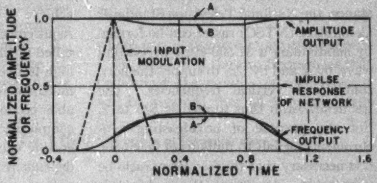


Fig. 18. Approximation to output amplitude and frequency response of network with rectangular impulse response to F-M signal with triangular modulation of one-quarter width of impulse response

A—Zero peak deviation
B—One-half bandwidth

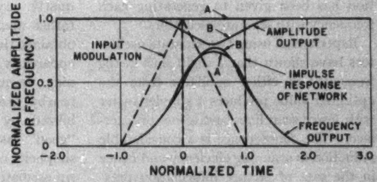


Fig. 19. Similar to Fig. 18 except that modulation has same width as impulse response

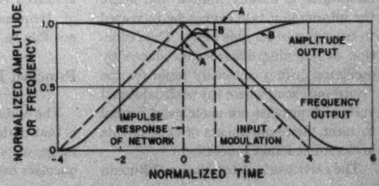


Fig. 20. Similar to Fig. 18 excpt that modulation is four times width of impulse response

3. THEORIE ET APPLICATIONS DE LA NOTION DE SIGNAL ANALYTIQUE (Theory and Applications of the Concept of the Analytic Signal), J. Ville. *Cables et Transmission*, Paris, France, vol. 2, 1948, p. 61 et seq.

4. THE THEORY OF ANALYTIC BAND LIMITED SIGNALS APPLIED TO CARRIER SYSTEMS, J. R. V. Oswald. *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., vol. CT-3, 1956, p. 244 et seq.

5. INTRODUCTION TO THE THEORY OF FOURIER INTEGRALS, E. C. Titchmarsh. Oxford University Press, London, England, 1937.

6. THE TRANSMISSION OF A FREQUENCY MODULATED WAVE THROUGH A NETWORK, W. J. Frantz. *Proceedings*, Institute of Radio Engineers, vol. 34, 1946, p. 114 et seq.

7. INTERCHANNEL INTERFERENCE IN FM AND PM SYSTEMS UNDER NOISE LOADING CONDITIONS, W. R. Bennett, H. E. Curtis, S. O. Rice. *Bell System Technical Journal*, vol. 34, 1955, p. 601 et seq.

8. DISTORTION OF FREQUENCY MODULATED SIGNALS IN ELECTRICAL NETWORKS, F. L. H. M. Stumpers. *Communication News*, New York, N. Y., vol. 10, 1948, p. 82 et seq.

9. VERZERRUNGEN IN FREQUENZMODULIERTEN SYSTEMEN (Distortion in Frequency-Modulated Systems), A. Ditl. *Hochfrequenztechnik und Elektroakustik*, Leipzig, Germany, vol. 65, 1957, p. 136 et seq.

10. EINSCHWINGVORGANGE BEI FREQUENZ-MODULATION (Step Functions in Frequency Modulation), A. Ditl. *Ibid.*, vol. 64, 1956, p. 184 et seq.

Two-Oscillator Scan Stability

HORACE M. JOSEPH
MEMBER AIEE

THE USUAL regularly repeated raster scan, such as used for television picture presentation, is the result of two perpendicular deflection waves which produce a complex Lissajous figure in the form of a set of closely interwoven lines. The number of lines that appear in this raster during each repetition is an important factor in determining the amount of video information that can be presented.

Ordinarily both of the (different frequency) waveforms are derived from a common sine-wave signal by generating two appropriate submultiples. For instance, the National Television Standards Committee (NTSC) raster can be formed by subdividing a 31,500-cycle-per-second wave by 2 and by 525 to supply the horizontal and vertical frequencies. The frequency ratio thus formed is 525 to 2. For the purpose of broadcasting synchronizing signals to millions of receivers, it is necessary that the frequency ratio be very stable. Apparently the need for this great stability is the reason little attention has been given to generating each frequency with separate oscillators.

Experiments using two separate oscillators have shown that a slight shift in the frequency of either oscillator from the optimum value produces a simpler raster with fewer scan lines and a loss of interlace. The consequence is a considerable reduction in scanning efficiency and hence in the rate of picture information presentation. The available video-circuit bandwidth is usually limited and hence the information rate cannot be increased to compensate for this loss. The reduction of scanning efficiency produced by oscillator drift is especially rapid for the "isotropic" scan where the horizontal and vertical frquencies are made nearly equal to meet the requirements of experiments in electronic image processing.

The early work at the National Bureau of Standards[1] was done with an isotropic scan that had no interlace. To avoid possible instability, synchronization was obtained by using submultiples of a common frequency. The frame rate was low, about one-half frame per second. Later the Avion Division–Alexandria built an image-processing equipment for use by the Air Force RADC. This equipment required conventional frame frequencies and, hence, higher scanning frequencies. To remove flicker effects, and also minimize video bandwidth, necessitated a quadruple interlace. To achieve a raster scan that met those requirements from a common signal entailed much circuitry for subdivision of a high-frequency and accurate frequency multiplication. Consequently, appreciable economy and reduction of circuit complexity could be obtained by generating each scan frequency with an independent oscillator.

It was recognized that each oscillator should be stabilized by a piezoelectric quartz crystal. This was done, but rather unexpectedly the raster stability obtained this way was insufficient for some apparently properly chosen pairs of scanning frequencies. Contrariwise, other nearby proper pairs of frequencies, stabilized by similar crystals, provided "good" rasters. In view of the circuit-simplicity obtained by using two independent oscillators and the demonstrated possibility of obtaining favorable results, it appeared valuable to determine the range of frequency variation that made it possible to obtain good scanning rasters.

Principal Raster Effects

The stability problem considered here is that of determining the largest relative range of one of the two deflecting frequencies over which the number of raster lines remains equal to or larger than the required number. It might be looked upon as the frequency range that maintains the complexity of a Lissajous figure above a given level. Once a frequency ratio has been determined, as given later, the sole cause of raster changes is a drift of scanning ratio, produced by a change in the frequency of one or both oscillators. Although many writers[2] have discussed the factors that affect the frequency of individual oscillators, little indication has been given of the raster effects produced by relative frequency drift. In addition, there are other factors that affect the number of raster lines observed by a viewer.

DRIFT OF SCANNING RATIOS

Fig. 1 shows a simple raster formed by two deflecting waves whose frequencies are nearly alike. Very little drift in frequency is required to alter the number of times the scanning spot traverses the raster before it returns to its starting point. Because of the integral numbers involved, the effect of frequency changes is determined by the properties of integers, those obtained by expressing the ratio of the number of vertical lines to the number of horizontal lines as a fraction reduced to lowest terms. An example of the rapidity with which the members of a "reduced" fraction can change as a result of a relatively small change in its numerator is 29/60. It becomes 1/2 (=30/60) by adding 1 to 29 and thus, when reduced, changes its denominator drastically. Since the ratio of uncontrolled oscillator frequencies can vary continuously, fractional changes in the numerator can occur, e.g., the numerator of 20/81 can drift to a somewhat larger value to produce $1/4(=20\frac{1}{4}/81)$. It will be shown that the permissible

---

Paper 59-1179, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted June 18, 1958; made available for printing August 28, 1959.

HORACE M. JOSEPH, formerly with Avion Division, ACF Industries, Alexandria, Va., is now with The Johns Hopkins University, Silver Spring, Md.

Vol.
Pt. 1
195

range of drift can be found from the solution to the problem of determining the numerator range over which the lowest terms denominator of the fraction remains large enough. (The finite observation time prevents the occurrence of an infinite denominator.) Illustrations of the effect of drift in denominator value are given in Fig. 2.

PHASE

The position relation or relative phase difference between a point of the vertical y-wave and a corresponding point on the horizontal x-wave affects the number of interlaces. This change is one that causes some television sets to present black spaces between scan lines. It is shown in Appendix I that, in general, a drift in phase angle causes a drift from half the maximum number of raster lines up to the maximum number and down to half, etc. Fig. 3 presents an example of the effect of variation in phase angle upon the raster of Fig. 1.

TIME OF OBSERVATION

The preceding considerations are based on frequency ratios, which are essentially mathematical abstractions. The scanning frequencies are never exact numbers, but are approximations. The scan lines in succeeding frames do not exactly superimpose, but they appear to do so when they are "close enough." Also, the apparent number of raster lines depends upon the time of observation, for instance, the perception period of a viewer or the length of a photographic exposure. This time is usually short and, consequently, it is necessary to consider the ways of obtaining optimum, fine-structured raster coverage in order to scan most efficiently and gain as much image information as possible during the available time. Fig. 4 illustrates the loss of scanning efficiency that occurs on a television screen when the effective raster coverage is reduced (on one half).

TOTAL PERCEPTIBLE SCREEN RESOLUTION

The number of scan lines is also limited in practice by the number of scan lines that can be resolved either on the screen or by the viewing means.

DEFLECTION WAVEFORM

The waveform of each deflecting wave has been assumed to be one used for scanning. However, the calculations that follow represent counts of the number of lines and do not assume any distribution. It follows that the results apply to any "simple" waveform such as a sawtooth, sinusoid, or triangle, where each deflection wave proceeds monotonically from one extreme to the other and also returns monotonically, but possibly in a different way. In addition, the horizontal and vertical waveforms are assumed alike, except in period and amplitude.

For ease of illustration and visualization the deflection waveforms will be drawn as if they were symmetrical triangles. To avoid ambiguity, the phase will be expressed as a fraction of a wavelength or period rather than as an angle.

Relative Oscillator Stability

The calculation of the required relative oscillator stability is made by determining the ideal, or infinite, observation time, "permissible" drift of one scanning frequency relative to the other, and then finding how near a "practical" frequency can approach this value without departing from the conditions of maintaining the minimum usable number of scan lines. The scanning frequencies themselves are initially chosen on the premise that each should be as low as possible to reduce scanning velocity in order to minimize the effects of phosphor persistence and to reduce the bandwidth needed to reproduce picture detail with the least effect of random noise.[8] Consequently, the permissible change in scan line number must be very small. The value chosen for the change in the larger number of x-waves is unity and no decrease is permitted in the number of interlaces. The number of scan lines in a frame (for constant scanning frequencies) is determined by the number of x-wave and y-waves that exist between a particular pair of position-reference values and the reoccurrence of that pair. For convenience of discussion, the y-deflection period is chosen to be the longer period. It can then be related to the x-deflection period as follows:

$$T_y = T_x(1+a) \qquad (1)$$

where a is a positive quantity. Since it is the relation of the two frequencies that determines the number of scan lines, it will be assumed that all frequency drift

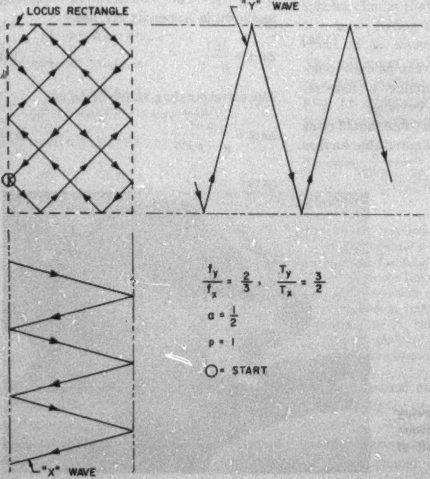
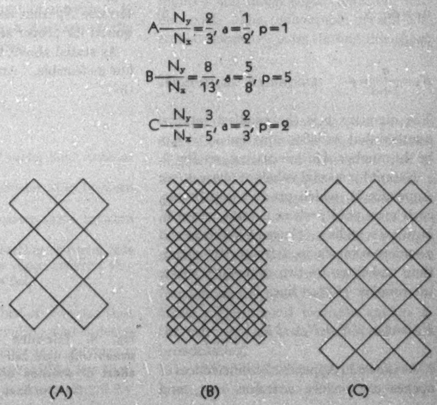

LOCUS RECTANGLE
"Y" WAVE
$\frac{f_y}{f_x} = \frac{2}{3}, \quad \frac{T_y}{T_x} = \frac{3}{2}$
$a = \frac{1}{2}$
$p = 1$
○ = START
"X" WAVE

Fig. 1 (left). Example of triangular wave scan

Fig. 2 (below). Effect of change in scan ratio


A—$\frac{N_y}{N_x} = \frac{2}{3}, \; a = \frac{1}{2}, \; p = 1$
B—$\frac{N_y}{N_x} = \frac{8}{13}, \; a = \frac{5}{8}, \; p = 5$
C—$\frac{N_y}{N_x} = \frac{3}{5}, \; a = \frac{2}{3}, \; p = 2$
(A)
(B)
(C)

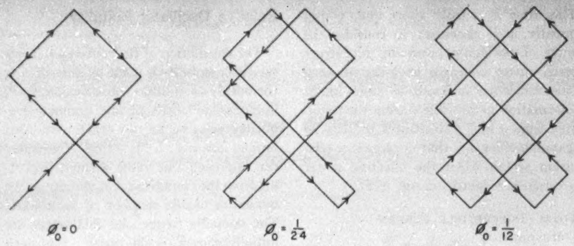


Fig. 3. Examples of effect of phase change in Fig. 2(A)

is produced by a change in the value of a. After a time, t, elapsed after the first time $x=0$, the number of x-wave periods is (an integer plus a fraction)

$$\phi_x = \frac{t}{T_x} \tag{2}$$

Also, the number of y-wave periods is (an integer plus a fraction)

$$\phi_y = \frac{\phi_x}{1+a} + \phi_0 \tag{3}$$

where ϕ_0 is the portion of the y-wave existing when $t=0$. The difference of the two phases is

$$\Delta\phi = \phi_y - \phi_x = \phi_0 - \frac{a}{1+a}\,\phi_x \tag{4}$$

When the time t is equal to the frame time, everything begins again. Consequently, at the end of a frame, an integral number of x-waves has occurred and also a different number of y-waves have passed. Then

$$\phi_x = N_x \tag{5}$$

an integer, and $\Delta\phi$ differs from ϕ_0 by another integer p, that is

$$\Delta\phi = \phi_0 - p \tag{6}$$

Hence

$$N_x = \frac{a}{1+a} = p \tag{7}$$

The quantity p is the smallest integral number that satisfies equation 7; it can be the number of interlaces, e.g., see Fig. 2.

Except for special values of phase ϕ (see Appendix I), which produce repetitions, each apex point, where a wave reverses, begins a scan line. Since no x-apexes and y-apexes coincide in a frame (by definition) and there are two apexes in a wave, the number of scan lines is

$$N_s(\max) = 2(N_x + N_y) = 2N_x\left(\frac{2+a}{1+a}\right) \tag{8}$$

As shown in Appendix I, coincidences of apexes can reduce a frame once, and and then only if ϕ_0 has an appropriate value. When an appropriate value of ϕ_0 occurs, there will be a reduction of the number of scan lines to a minimum of

$$N_s(\min) = \frac{2+a}{1+a}\,N_x - 1 \tag{9}$$

The average number of scan lines depends upon the way ϕ_0 varies and upon the value of the minimum separation that can be perceived between adjacent lines; see Appendix II. For typical conditions of slow, uniform frequency drift and closely spaced lines, the average number of lines is the arithmetic average of equations 8 and 9,

$$N_s(\text{avg}) = \frac{3}{2}\,N_x\,\frac{2+a}{1+a} - \frac{1}{2} \tag{10}$$

Under the conditions of nearly equal scanning rates, $N_x >> 1$, and $a << 1$,

$$N_s(\max) = 4N_x \tag{8A}$$

$$N_s(\min) = 2N_x \tag{9A}$$

and

$$N_s(\text{avg}) = 3N_x \tag{10A}$$

Equations 8, 9, and 10 can be used to determine N_x from the number of lines required for proper scan coverage.

As stated above, this value should be as low as feasible. Any appreciable further reduction in its value is considered intolerable. Hence the permissible value of

$$\Delta N_x = \pm 1 \tag{11}$$

or

$$\Delta p = 0 \tag{12}$$

whichever occurs sooner. The latter condition is more stringent if $p>1$. An example of the raster deterioration that results from exceeding the limit is diagrammed in Fig. 2. This follows from the Diophantine nature of equation 7, which is satisfied only with the smallest integral numbers for N_x and p. Relatively small variations, Δa's, can produce large changes in p and consequently in N_x.

The reduction can be expressed algebraically, and the values of Δa required to produce the change determined. Two different values of Δa can be found; a "conservative" value calculated relatively easily when the number of lines, N_s, is not very large, and a wider range of Δa for use when the required frequency stability should be minimized. The calculation begins by writing equation 7 in a different form:

$$\frac{N_x}{p} = M_x + \frac{n}{p} \tag{13}$$

where

M_x = the integral number of waves in a field
n = an integer less than p

Then determine the positive and negative changes in n, Δn_+, and Δn_- (not necessarily integers), which make p drop to $p-1$. The conservative answers, as shown in Appendix III, are

$$\Delta n\pm = \frac{\pm 1}{p-1} \tag{14}$$

The corresponding values of Δa are

$$\Delta a\pm = \frac{\pm a^2}{p^2 - p \pm a} \tag{15}$$

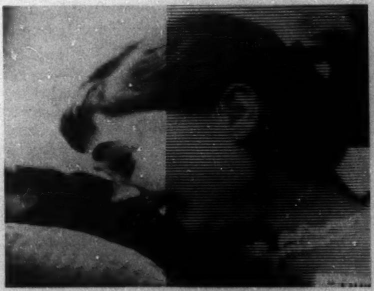


Fig. 4. Television viewing screen with right half showing effect of omission of half of the scan lines

The total range of Δa is

$$\Delta a = \frac{2(p^2-p)a^2}{(p^2-p)^2-a^2} \tag{16}$$

For the important case of no interlace, $p=1$, equation 11 is used. Then

$$\Delta a \pm = \frac{\mp a^2}{1+a} \tag{15A}$$

For the small a's the over-all range of Δa becomes

$$\Delta a \doteq 2a^2 \tag{15B}$$

A larger range, shown in Appendix III, is possible. For instance the range can be increased when $n=2$. Also still larger values are possible, when n is larger, if the numerator and denominator of the fractions,

$$\frac{n-1}{p-1}, \frac{n-2}{p-2}, \ldots$$

do not contain a common factor; each specific value should be tested. The largest possible ranges of Δn that make p drop to $(p-1)$ are

$$\Delta n_{+} = \frac{n}{p-1} \tag{17}$$

and

$$\Delta n_{-} = \frac{n-1}{p-1} \tag{18}$$

The corresponding values of Δa that will produce these Δn's are

$$\Delta a_{+} = -\frac{a^2 n}{p^2-p+an} \tag{19}$$

and

$$\Delta a_{-} = \frac{a^2(p-n)}{p^2-p-a(p-n)} \tag{20}$$

The over-all range is

$$\Delta a = \frac{a^2 p^2(p-1)}{(p^2-p+an)(p^2-p-ap+an)} \tag{21}$$

For small a's this reduces to

$$\Delta a \doteq \frac{a}{p-1} \tag{21A}$$

This value of Δa is larger than that indicated by equation 16, but caution is required. These equations show that the permissible oscillator drift decreases when the interlace is increased for a given ratio, a.

PRACTICAL CONSIDERATIONS

The stability considerations are most critical when a is small and, consequently, when equations 15(B) or 21(A) apply. Further, it is usual for $N_x >> p$. Therefore the value of n makes little difference in scanning velocity. However, the particular choice made may create an important difference in the rate of raster deterioration. The example of Fig. 2 was chosen to show a rapid change. The quantity n should be chosen so that n and $p-1$ are relative primes; if possible, the other conditions for equations 17 and 18 should be met. It was failure to meet this condition that produced the unsatisfactory pairs of crystal-controlled oscillator frequencies where the raster was not stable enough.

PRACTICAL FREQUENCY

The above calculated frequency deviation ratios are based on the properties of integers. They are abstractions which which can only be approached. Because of a finite observation time, such approaches are seen; for example, many people look at patterns in the form of television rasters or of Lissajous patterns on oscilloscopes. This closeness of approach depends on the minimum perceptible separation between scan lines perceived during the "time of observation." These quantities are functions of observer and equipment characteristics. Their values determine how close an approach to the calculated ends of the frequency ranges can occur without perceiving raster deterioration. The closeness of approach can be used as corrections derived from the measured perceptible line separation distances and observation times, by using the idea of a practical scanning frequency.

This frequency can be determined from the minimum average perceptible line spacing change that occurs in the observation time, taking account of the average effect of ϕ_0 discussed in Appendix II. The minimum visible spacing is defined as a fraction of the length of one side. It may be expressed as a fraction of the separation between lines:

$$S_v = vs = \frac{v}{2N_x} = v\frac{\theta}{2p} \tag{22}$$

where

S_v = visible fraction of raster side required for visibility
s = average fraction of raster side between adjacent lines
v = visible fraction of raster space between adjacent lines
θ = ratio p to N, the fraction of raster side occupied by the space between successive lines in a field

The practical number of x-cycles that occur during the observation time is designated N_{Tx}. With any change in spacings, a result of variation in a should be less than S_v. This change in a can be found by solving for Δa in an equation derived from equation 5:

$$\Delta\phi - \Delta\phi_T = \phi_0 - \frac{a}{1+a}N_{Tx} - \left(\phi_0 - \frac{a+\Delta a}{1+a+\Delta a}N_{Tx}\right) = \pm v\frac{\theta}{2p} \tag{23}$$

The resulting correction in Δa is used to reduce the variations permitted by equations 19 and 20. Inserting the correction reduces Δa_{+} to

$$\Delta a_{+}' = \frac{-a^2(1+a)\left[n\frac{N_{Tx}}{N_x} - \frac{p-1}{p'}\frac{v}{2}\right]}{(p^2-p+an)\times\left[(1+a)\frac{N_{Tx}}{N_x} - \frac{a^2}{p^2-p+an}\frac{v}{2}\frac{p-1}{p'}\right]} \tag{24}$$

where

p' = interlace that the first permissible positive increase in n may produce

The resulting corrected Δa_{-} is

$$\Delta a_{-}' = \frac{a^2(1+a)\left[(p-n)\frac{N_{Tx}}{N_x} - \frac{p-1}{p''}\times\frac{v}{2}\right]}{[p^2-p-a(p-n)]\times\left[(1+a)\frac{N_{Tx}}{N_x} + \frac{a^2}{p^2-p-a(p-n)}\times\frac{v}{2}\times\frac{p-1}{p''}\right]} \tag{25}$$

where

p'' = the reduced interlace that the permissible decrease in n may produce

Note that under relatively favorable conditions when N_x and n are chosen as so to obtain the proper relative primes (see Appendix III) either $p'=p-1$, or $p''=p-1$; if they are most favorable, $p'=p''=p-1$. However, when a good choice is not made, the permissible variation is reduced and under some conditions may not exist. The unfavorable crystal-controlled frequency pairs showed this effect.

Numerical Examples

Two numerical examples are given in the following to indicate the application of the equations to determine the 2-oscillator stability required to produce a good raster coverage with the number of lines called for by the conventional NTSC standard. The first example applies the calculations to an isotropic scan where the vertical and horizontal scan frequencies are nearly alike; where a is small. The second example applies to the conventional scan used in television broadcasting.

1. In addition to $a<<1$, and $N_x=525$ lines, $p=5$ is chosen. It is assumed that

the phosphor material and viewing screen size are chosen so that

$$\frac{N_{Tx}}{N_x}=10$$

and

$$v=0.5$$

From equation 10(A)

$$N_x=175 \text{ cycles}$$

From equation 7

$$\frac{175}{5}=35$$

but 175 is divisible without remainder by 5 and in that event $p=1$. Since this is not acceptable, N_x is increased to 176 It is found that $p'=2$ and $p''=1$. This produces a narrow range of Δa; hence N_x is increased to 177. For this choice

$$p'=p''=p-1=4$$

From equation 7,

$$a=\frac{5}{172}$$

From equations 24 and 25 the combined range is

$$\Delta a'=\Delta a_-'-\Delta a_+'=2.5\times 10^{-4}$$

However, after a warm-up period, the frequency of an oscillator wanders randomly around a central value. The probability of reaching the limit given above should be reduced to a low value, say 0.1%. Hence the corresponding 50% probability value[4] of the necessary oscillator stability is

$$\pm\frac{\Delta a}{2}=\pm 0.005\%=\pm 1/20{,}000$$

2. It may be of some interest to note that under the same conditions of observation time and perceptible line separation the relative oscillator stability required for the conventional NTSC scan of 525 horizontal and vertical scans per frame can be calculated. The corresponding quantities are

$N_x=525$ cycles

$a=523/2$

$N_s=1{,}054$ (The effective value is reduced by blanking.)

$p=523$ (The total number of horizontal and vertical cycles that are interlaced.)

From equation 13

$$\frac{N_x}{p}=2+\frac{4}{523}$$

and

$$n=4$$

Searching through the various possibilities of

$$\frac{1}{p-1},\ \frac{1}{p-2},\ \ldots$$

Appendix III shows that equations 24 and 25 apply with $p'=p''=p-1=522$. Accordingly

$$\Delta a_+'=-0.9919$$

and

$$\Delta a_-'=258.5$$

This indicates that N_x may be reduced by very nearly a whole cycle or more than doubled without causing a decrease in the number of raster lines as a result of changer in the ratio of scan frequencies, except for interlace changes due to phase-angle change.

Conclusions

The examples and formulas for the calculation of the relative stability required of the oscillators in the 2-oscillator method of producing a scan, when due regard to providing for phase-angle variation is given, provide stability tests for choosing frequency pairs. The results show that the frequency stability required can be well within that obtainable from piezoelectric crystal-stabilized oscillators. For exacting conditions when the viewing time is short, some searching is required to find frequencies that will provide a maximum range of variation. Although this must be done by arithmetic trials, the procedure for checking the range of stability is readily done.

The results also show that schemes which require nearly equal (small a) frequencies requires more stability. Consequently the number of raster lines may be very sensitive to relatively small influences that change the effective scan frequency, such as hum or noise frequency modulation of either oscillator, or an equivalent scan frequency modulation caused, for example, by variation of electron beam stiffness resulting from changes in accelerating voltage.

The calculations, except for the interpretation of phase, do not depend upon the waveform of the scan signals. They apply to sinusoidal Lissajous figures as well as those produced by sawtooth or symmetrical triangular waves. Consequently the formulas may be useful in the measurement of frequency by oscilloscope comparison, using a Lissajous figure.

Appendix I. Effect of Phase Angle

The initial value of the y-wave at $t=0$ is the phase fraction ϕ_0. Its value affects the number of scanning lines that occur in a frame as stated in equations 8, 9, and 10. The proof of these equations is completed by examining the behavior of $\Delta\phi$. (To facilitate visualization the proof is given for a symmetrical triangular wave. Neither the symmetry nor triangularity is essential.) See Fig. 5. First assume that

$$\phi_0=0$$

Then at the beginning of an x-wave, where $x=0$, equation 4 shows that

$$\phi_y=\frac{m_x}{2}+\Delta\phi$$

where

m_x = (integral) number of half x-waves that occurred after the beginning of the frame

The y-wave will reach its apex when

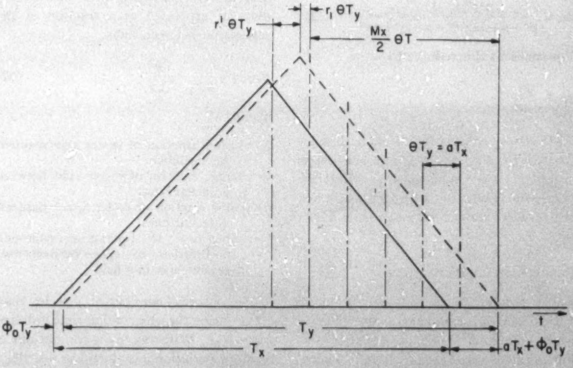


Fig. 5. Diagram of relation between x-waves and y-waves

$$\phi_y = \frac{m_x}{2} - \frac{1}{2}$$

and again reach zero when

$$\phi_y = \frac{m_x}{2} - 1$$

Also from equation 4, it is apparent there is a retrogression, of magnitude, θ, in the y-wave period for each cycle of the x-wave. The y-wave will lag by a half-period when

$$\Delta\phi = \frac{1}{2}$$

Also

$$\theta = \frac{a}{1+a}$$

The half-period lag occurs when

$$N_x = \frac{m_x}{2} + \frac{n}{2p} \tag{26}$$

An alternative statement of this is that after $m_x/2$ complete periods of the x-wave, the y-wave lags by $r_1\,\theta$ less than 1/2 cycle, where

$$r_1 = \frac{n}{2p} \tag{27}$$

At the end of the next succeeding x-wave, the y-wave lags by 1/2 cycle plus a fraction,

$$r_1'\theta$$

and

$$r_1' = 1 - \frac{n}{2p} \tag{28}$$

If the scan lines are not to overlap, then

$$1 - \frac{n}{2p} \neq \frac{n}{2p}$$

or

$$n \neq p \tag{29}$$

This condition is met for all values of p except when $p=1$. The process continues for an additional $(m_x/2-1)$ waves if

$$2n < p$$

or for $\frac{m_x}{2}$ waves if

$$2n \geq p$$

until $\Delta\phi$ is short of -1 by a fraction r_2 of θ, and

$$r_2 = 1 + \frac{n}{2p} - r_1' = \frac{n}{p} \tag{30}$$

In general, for the kth apex the fraction

$$r_k = 1 + \frac{n}{2p} - \left[1 - \frac{(k-1)n}{2p}\right] - m = \frac{kn}{2p} - m \tag{31}$$

and another fraction

$$r_k' = 1 - r_k = 1 + m - \frac{kn}{2p} \tag{32}$$

where m is the integer required to make r_k and r_k' proper fractions.

Since

$$1 \leq k < 2p \tag{33}$$

it follows that when

$$r_k = r_k'$$

$$\frac{kn}{p} = 1 + 2m \tag{34}$$

This, of course, can only occur when

$$k = p \tag{35}$$

since the fraction n/p is reduced to lowest terms, and then can occur only if n is an odd integer.

There may be additional coincidences. These may happen when k is odd near the crest, and the other when k is even. Substituting another index i for k and j for m, another set of r's is

$$r_i = \frac{in}{2p} - j \tag{36}$$

and

$$r_i' = 1 + j - \frac{in}{2p} \tag{37}$$

From the definition of p, when $i \neq k$

$$r_i \neq r_k \tag{38}$$

but consider the possibility that

$$r_i' = r_k$$

Then from equations 33, 36, 37, and 38

$$\left(\frac{k+i}{2}\right)\left(\frac{n}{p}\right) = 2m + j + 1 \tag{39}$$

Since

$$\frac{k+i}{2} < 2p \tag{40}$$

only one set of values of $(k+i)/2$, namely opposite ones, can make equation 39 correct. There is, therefore, one, and only one, set of overlapping lines. By similar reasoning when

$$\phi_0 \neq 0$$

$$\left(\frac{i+k}{2}\right)\left(\frac{n}{p}\right) = m + j - 2r_0 + 1 \tag{41}$$

where

$$r_0 = \frac{\phi_0}{\theta} \tag{42}$$

This shows that again only one overlap can occur; however, it occurs only if r_0 is an integral multiple of $n/2p$.

Appendix II. "Perceptible" Effect of Phase Angle

Equation 41 expresses another numerical-algebraic result which must be converted into practical terms of perceptible line spacing Now compare equation 22 which is valid when equation 41 is not satisfied with the equation when equation 41 is satisfied. The separation between adjacent lines is then

$$s_1 = \frac{\theta}{p} = 2s \tag{43}$$

where

s = separation given in equation 22

If r_0 progresses uniformly in value from one integral multiple of $n/2$p to the next, then s_1 decreases uniformly to s and increases again uniformly to $2s$. Because of this uniformity in rate the fractional part of the total time of that progression, when equation 43 effectively applies, is also v. The average number of lines, N_s, is the average of equations 8 and 9 weighted by v. Hence

$$N_s(\text{avg}) = N_x \frac{2+a}{1+a}(2-v) - v \tag{44}$$

When $a \ll 1$, and $N_x \gg 1$,

$$N_s(\text{avg}) \doteq 2(2-v)N_x \tag{44A}$$

Equation 44 becomes equation 10 when $v=1/2$, a usual value.

Appendix III. Obtaining a Sufficiently Large Denominator

There are four conditions which limit the range of Δn within which a fraction whose denominator remains equal to or larger than p over that range can be obtained. These conditions are discussed in the following in terms of a Δn expressed as a rational fraction, reduced to lowest terms, added to n.

1. The resulting new fraction is

$$\frac{n + \frac{\alpha}{b}}{p} = \frac{bn+\alpha}{pb} \tag{45}$$

where α and b are integers. Now the numerator $(bx+\alpha)$ is not factorable without remainder by b. However, it is possible that some choices of b and α will make the numerator factorable by p without remainder. In that event the denominator will be b. In order that the deterioration of the interlace and consequent number of scanning lines be limited, it is necessary to make

$$b \geq (p-1) \tag{46}$$

2. The fraction α/b can have any magnitude since a can vary continuously. This means that α should be limited to ≤ 1, since any fraction less than $1/b$ will have a denominator larger than b. This can be shown as follows:

By hypothesis that fraction

$$\frac{A}{B} < \frac{1}{b} \tag{47}$$

where A and B are positive integers.

From that

$$B > Ab$$

Hence

$$B > b \tag{48}$$

since A is an integer.

These two conditions are relatively easy to check. Consequently, they are the basis of the "conservative" values of Δa in equations 15 and 16.

3. On the other hand, to use a value of α larger than 1 may permit a reduction (if the numerator can be divided by p without remainder) of b below $(p-1)$. There is a risk, e.g., if $\alpha=2$ and b is an even number, the effective value of the denominator is $b/2$. However, the numerator is not always divisible by p when $b=(p-1)$. In that case, it is possible that the denominator is greater than p until, possibly, b is reduced to $(p-2)$, and if not then, Δn may be increased to $1/(p-3)$, and so on. Let

$$\Delta n=\frac{1}{p-d} \tag{49}$$

where

d = an integer less than p

Then

$$\frac{n+\Delta n}{p}=\frac{n-\dfrac{dn-1}{p}}{p-d}$$

The numerator will not be an integer unless

$$\frac{dn-1}{p}=\text{an integer} \tag{50}$$

However, if it is not an integer for a particular value of b, there is still the possibility that the fraction can be reduced to lower terms. Since the denominator of that fraction cannot be smaller than 2, to be safe, make

$$d<\frac{p}{2} \tag{51}$$

4. The limit of variation Δn is

$$\Delta n_{+}=\frac{n}{p-1} \tag{52}$$

since the reduction of the denominator to $(p-1)$ will then occur, and provided

$$\frac{1}{p-d}>\frac{n}{p-1}$$

for the first d that satisfies equation 50. Similarly, the limit of negative variation of Δn is

$$\Delta n_{-}=\frac{n-1}{p-1} \tag{53}$$

provided:

$$\frac{1}{p-d}>\frac{n-1}{p-1}$$

as above. It is also plain that equations 52 and 53 always apply to the case where

$n=1$, or 2

since the new denominator will be $p(p-1)/2$ $(>p)$ for $n=2$ (this condition also means that p is an odd number).

References

1. Image Processing, L. S. G. Kovasznay, H. M. Joseph. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 43, no. 5, May 1955, pp. 560–70.

2. Radio Engineering (book), F. E. Terman. McGraw-Hill Book Company, Inc., New York, N. Y., 3rd edition, 1947, p. 417.

3. Television (book), V. K. Zworykin, G. A. Morton. John Wiley & Sons, Inc., New York, N. Y., 2nd edition, 1954, p. 194.

4. Handbook of Physics and Chemistry (book), C. O. Hodgmond et al. Chemical Rubber Company, Cleveland, Ohio, 35th edition, p. 229.

Electronic Network Synthesis of Linear Algebraic Matrix Equations

ROBERT E. HORN
ASSOCIATE MEMBER AIEE

PIERRE M. HONNELL
MEMBER AIEE

APPLICATIONS of matrix techniques to the electronic network synthesis of several types of equations have been discussed in previous publications.[1–5] A brief review of the basic theory is included in the Appendix. The objective of this paper is to apply these techniques to the synthesis of linear algebraic matrix equations. Existing methods are summarized and their domain of usefulness is indicated, particularly with respect to network stability.

Goldberg and Brown[6] and Korn,[7] emphasize that a paramount problem in solving algebraic systems on electronic network machines or computers is, surprisingly, that of stability; this is in contrast to the application of such electronic computers to the solution of differential equations systems, in which case stable synthesis networks are easily devised.

An ingenious solution to the stability problem was described by Adcock[8] on an electromechanical machine computer; the same principle was applied to an all-electronic computer by Gephart.[9] These apparently diverse methods are easily demonstrated, by a matrix formulation, to be closely akin.

The central problem is that of exhibiting the solution of the algebraic system

$$Ax=b \tag{1}$$

where $[A]$ is a square nonsingular matrix of real constant elements, x is a column matrix of n unknowns to be determined, and b is a column matrix of n prescribed real elements. This systems may also be represented as:

$$\sum_{j=1}^{n} a_{ij}x_j=b_i, \quad i=1, 2, \ldots, n \tag{2}$$

In the ideal electronic network machine computer, there would arise no stability problem, the determinant of $[A]$ being nonvanishing, since the system is algebraic. The synthesized network in an electronic computer is, however, physical; as such it then represents a differential equation, with the inherent stability problem with respect to oscillations. This arises from the inevitable deviations of the electronic amplifiers from the ideal, that is, a voltage amplification which is a positive (or negative) real constant; and also from parasitic effects, principally capacitive. Both of these factors introduce differential coefficients into the system; hence the differential equation.

Since it is not possible to synthesize a purely algebraic equation when high-amplification electronic amplifiers are involved, the key concept arises of converting the given mathematical algebraic problem to a mathematical differential equation system which has as its particular integral the desired solution to the original algebraic system and which has a stable synthesis on an electronic computer. In this conversion of equation systems, mathematical stability alone is not sufficient to insure computer stability.

Stability Problem

It is imperative for clarity to distinguish between two types of stability problems. One type arises from the statement of the mathematical problem per se, here purposely introduced by differential equations as such. The other type is the true

---

Paper 59-1136, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted October 16, 1958; made available for printing August 25, 1959.

Robert E. Horn is with Westinghouse Electric Corporation, Baltimore, Md., and Pierre M. Honnell is with Washington University, St. Louis, Mo.

The authors acknowledge with pleasure the assistance of Donald F. Wann who obtained the oscillographic data presented in Figs. 2 and 4.

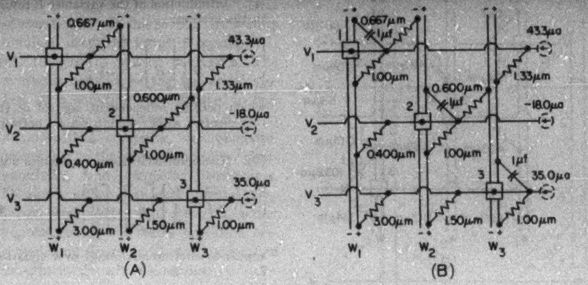


Fig. 1. Ideograph of numerical example

A—Using synthesis equation 1
B—Using synthesis equation 11

physical network stability that is actually encountered in the machine. The latter arises from a combination due to the intentionally introduced mathematical differential equation network representation and those arising from the parasitic parameters and amplifier characteristics.

The desired objective is to introduce the differential equation in such a manner that the stability is determined principally (entirely, if this could be achieved) by the stability consideration of the mathematical problem. In this paper the problem will be treated from the standpoint of systems known to be stable, rather than beginning with the generalized linear equations for mathematical stability and proving that nonideal amplifier characteristics do not introduce instability.

Adcock and Gephart introduce the following differential equation system, which is not only mathematically stable, but also network stable as shown by their experiments.

$$\sum_{j=1}^{n} A_{ij}x_j - b_i = e_i, \qquad i=1, 2, \ldots n$$

$$x_j = \sum_{i=1}^{n} E_i A_{ij}, \qquad j=1, 2, \ldots n$$

$$\frac{dE_i}{dt} = ke_i, \qquad i=1, 2, \ldots n \tag{3}$$

These consideration will be placed in a clearer light by the following: First of all, it is helpful to rewrite equation 3 in matrix notation as:

$$\begin{bmatrix} -I & A & 0 \\ 0 & I & {}_tA \\ -kI & 0 & D \end{bmatrix} \begin{bmatrix} e \\ X \\ E \end{bmatrix} = \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix} \tag{4}$$

where

$I = n$th order identity matrix
$D = I \cdot d/dt$
A, x, and B defined as in equations 1 and 2
e, E, and k defined as in equation 3
${}_tA =$ transpose of A

Equation 4 may be synthesized directly on an electronic machine and usually results in a network stable system if the matrix A itself is not positive definite. Premultiplication of matrix 4 by the transformation matrix

$$\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -I & 0 & \frac{1}{k}.I \end{bmatrix} \tag{5}$$

where I is the nth order identity matrix and k is positive real scalar, yields:

$$\begin{bmatrix} -I & A & 0 \\ 0 & I & {}_tA \\ 0 & -A & \frac{1}{k}D \end{bmatrix} \begin{bmatrix} e \\ x \\ E \end{bmatrix} = \begin{bmatrix} b \\ 0 \\ -b \end{bmatrix} \tag{6}$$

It is obvious that the desired solutions for x and E are obtainable from the reduced system

$$\begin{bmatrix} I & {}_tA \\ -A & \frac{1}{k}D \end{bmatrix} \begin{bmatrix} x \\ E \end{bmatrix} = \begin{bmatrix} 0 \\ -b \end{bmatrix} \tag{7}$$

in which network stability is further enhanced beyond that existing in the system 4. It should also be noted that if k is made indefinitely large, the form 7 goes over into that indicated by Goldberg and Brown.

The reason that the system 7 is likely to be more stable than 4 is not evident from these equations alone. The electronic network stability is determined by the complete network matrix which is the sum of the ideal synthesis matrix, the network error matrix, the equivalent source error matrix, and the parasitic error matrix.[8] In general, reducing the order of the system tends to reduce the effect of these error and parasitic matrices upon network stability.

The mathematical stability may be enhanced by a further reduction of the order of the system. Premultiplication of equation 7 by the transformation matrix of constant determinant

$$\begin{bmatrix} I & 0 \\ \frac{1}{k}D & -{}_tA \end{bmatrix} \tag{8}$$

results in

$$\begin{bmatrix} I & {}_tA \\ \left(\frac{1}{k}D + {}_tA \cdot A\right) & 0 \end{bmatrix} \begin{bmatrix} x \\ E \end{bmatrix} = \begin{bmatrix} 0 \\ {}_tA \cdot b \end{bmatrix} \tag{9}$$

Equation 9 as such is not the end in view, for it exhibits a zero matrix on the principal diagonal which, according to the previous discussion, emphasizes the undesired differential terms and may yield network instability. However, since this system is disjoint in a feedback sense, one can extract from equation 9 the following system:

$$\left(\frac{1}{k}D + {}_tA \cdot A\right)x = {}_tA \cdot b \tag{10}$$

Mathematically, equation 10 leads to a stable solution even if A itself is not positive definite. On the other hand, if the determinant of A is very small, then round-off errors become important; that is, the accuracy of the solution suffers from the finite resolution of the synthesis parameters. In fact, the squared determinant of A in equation 10 when

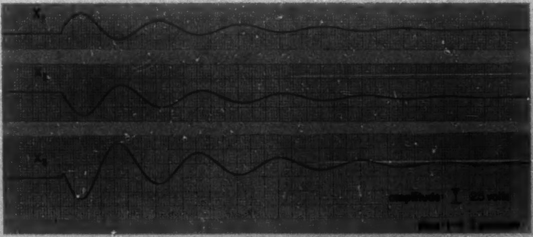


Fig. 2. Oscillogram corresponding to synthesis shown in Fig. 1(B)

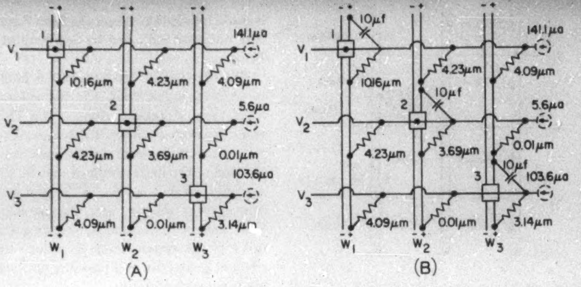


Fig. 3. Ideograph of numerical example

A—Using synthesis equation 12 B—Using synthesis equation 10

close to zero can as easily become negative as positive, due to the finite precision of setting of the synthesis network elements, and the computer thereby become unstable physically, vitiating the raison d'être of equation 10:

It should be emphasized that if A is known to be positive definite, the desired solution may be obtained directly from

$$\left(\frac{1}{k}D+A\right)x=b \tag{11}$$

without recourse to the forms 4, 7, or 10.

One obvious advantage of equation 11 over equation 10 is that in 10 the computation ${}_{t}A\cdot A$ and ${}_{t}A\cdot b$ must be performed by other means. In design applications this feature may be highly undesirable, because independent control of coefficients is lost. If the matrix A is positive definite even though almost singular, equation 11 is definitely superior to the others. Electronic computer application of these matrix equation forms may be summarized as follows:

1. Matrix A positive definite:

(*a*). $Ax=b$ (1)

may be utilized, but in many computers, inherent amplifier characteristics result in instability of the computer network approximation.

(*b*). $\left(\frac{1}{k}D+A\right)x=b$ (11)

results in the most usable form by masking amplifier characteristics, utilizing fewer amplifiers than equations 4 or 7 and maintaining independent control of coefficients.

(*c*). The other forms appear to offer no particular advantage in the application.

2. Matrix A not positive definite:

(*a*). ${}_{t}A\cdot Ax={}_{t}A\cdot b$ (12)

suffers the same limitations as equation 1 with the added disadvantage that the matrices ${}_{t}A\cdot A$ and ${}_{t}A\cdot b$ must be computed.

(*b*). $\left(\frac{1}{k}D+{}_{t}A\cdot A\right)x={}_{t}A\cdot b$ (10)

will yield stable computer networks, but as in equation 12, ${}_{t}A\cdot A$ and ${}_{t}A\cdot b$ must be computed.

(*c*). Introduction of the variables E results in the system:

$$\begin{bmatrix} I & {}_{t}A \\ -A & \frac{1}{k}D \end{bmatrix}\begin{bmatrix} x \\ E \end{bmatrix}=\begin{bmatrix} 0 \\ -b \end{bmatrix} \tag{7}$$

which permits independent control of the coefficients while maintaining network stability.

(*d*). The additional display variables e in

$$\begin{bmatrix} -I & A & 0 \\ 0 & I & {}_{t}A \\ -kI & 0 & D \end{bmatrix}\begin{bmatrix} e \\ x \\ E \end{bmatrix}=\begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix} \tag{4}$$

appear to offer no advantage over equation 7.

Experimental Results

A particularly good example of a critical system from the standpoint of electronic network stability is one furnished by Gephart:[9]

$$\begin{bmatrix} 1.00 & -0.667 & 1.33 \\ 0.400 & 1.00 & -0.600 \\ 3.00 & 1.50 & 100 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}=\begin{bmatrix} 43.3 \\ -18.0 \\ 35.0 \end{bmatrix} \tag{13}$$

The matrix has the following roots:

$$\begin{aligned} \lambda&=2.96 \\ \lambda&=0.01871\pm j0.238 \end{aligned} \tag{14}$$

These roots indicate that it should be possible to synthesize the network in all the forms previously indicated. Gephart gives the results obtained for the system in equation 4 and indicates also the use

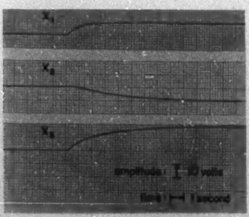


Fig. 4 (left). Oscillogram corresponding to synthesis shown in Fig. 3(B)

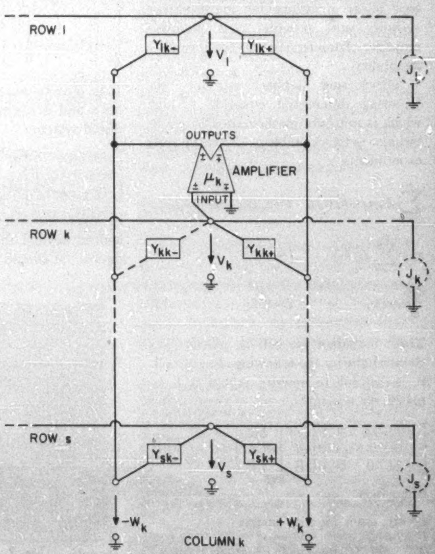


Fig. 5 (right). Electronic synthesis of kth column of mathematical matrix problem

of equation 7. In this section, the results obtained using the forms of equations 1, 11, 12, and 10 will be discussed. The synthesis network of the numerical system 13 in equation 1 is indicated in Fig. 1(A). When this network was placed on a commercial computer, instability resulted due to the inherent amplifier characteristics and parasitic parameters. However, utilizing an experimental computer, with simple *R-C* cutoff amplifier characteristics, the stable operation predicted from the roots 14 could be verified experimentally with the results:

$$x_1=10.1$$
$$x_2=-10.1$$
$$x_3=19.9 \qquad (15)$$

compared with the exact values

$$x_1=10$$
$$x_2=-10$$
$$x_3=20 \qquad (16)$$

The numerical equations 13 in the form

$$\left(\frac{1}{k}D+A\right)x=b \qquad (11)$$

result in the synthesis network depicted in Fig. 1(B). The case shown is for $k=1$. Obviously, the only distinction in the computing networks of Figs. 1(A) and (B) is in the addition of capacitors on each of the principal diagonal positions of Fig. 1(B). However, this form resulted in a stable system on the experimental and commercial computers, yielding the numerical values:

$$x_1=10.1$$
$$x_2=-10.1 \qquad (17)$$
$$x_3=20.0$$

The transient solution obtained is illustrated in the oscillogram of Fig. 2. These results indicate the advantage of equation 11 on commercial analog computers. By intentionally introducing the differential equation system, the parasitic terms tend to be masked and the actual network approaches more closely the ideal network. Of course, the solution response oscillates, as predicted by the roots 14 before settling down to the steady-state values.

Premultiplication of equation 13 by the transpose of the matrix of coefficients results in the equations:

$$\begin{bmatrix} 10.16 & 4.23 & 4.09 \\ 4.23 & 3.69 & 0.01 \\ 4.09 & 0.01 & 3.14 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 141.1 \\ 5.6 \\ 103.6 \end{bmatrix} \qquad (18)$$

and the corresponding roots of the characteristic equation for $k=0.1$:

$$\lambda=1.359$$
$$\lambda=0.340$$
$$\lambda=0.0000615 \qquad (19)$$

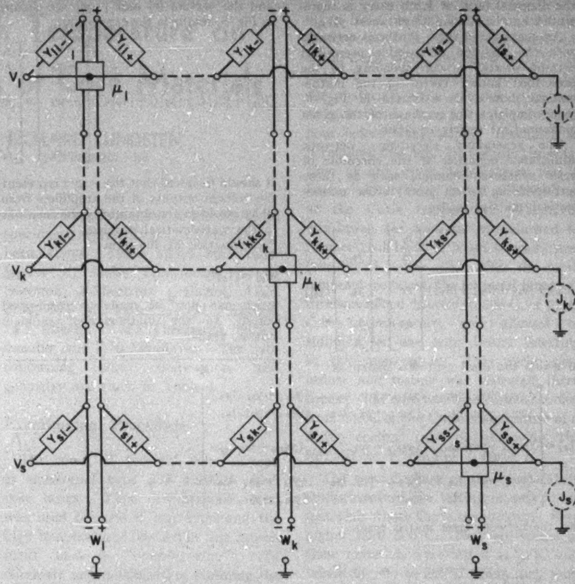


Fig. 6. Ideograph of ensemble of synthesis columns or complete synthesis network

The ideographs of Figs. 3(A) and (B) indicate the synthesis networks for the numerical equation 18 corresponding to equations 12 and 10 respectively. Again, the inclusion of the *d*:s on the principal diagonal in Fig. 3(B), was essential for stable operation of the experimental computer. Fig. 4 is the oscillographic record of the solution taken with $k=0.1$, and shows substantial agreement with the solutions previously obtained. As predicted by equation 19, the transient is now nonoscillatory.

The principal disadvantage of equation 18 as compared to equation 13 lies in the extreme dependence of the results of 18 upon the precision of the coefficient settings. In equation 18 changes of about 1% in the numerical values of the *b*:s, for example may cause errors of 100% or more in the measured magnitudes of the *x*:s, clearly indicating the desirability of avoiding this method if possible.

Conclusions

By use of the matrix formulation of the equation systems and the method of synthesis using the 1-to-1 correspondence between matrix equation and electronic network, a convenient tabulation of procedures applicable to solution of linear algebraic equations has been obtained. In addition, the use of the form

$$\left(\frac{1}{k}D+A\right)x=b \qquad (11)$$

has been shown to have considerable advantage in preventing instability resulting from inherent amplifier characteristics.

Appendix

The fundamental theory of the electronic synthesis of matrix equations is based upon an ensemble of the basic synthesis network shown in Fig. 5. It employs a combination of suitable amplifying devices such as electron tubes, transistors, and the like, and dinode network elements (such as capacitances, conductances, and inductances) or their combinations for the synthesis of one column of a mathematical matrix.

The synthesis of the *k*th column of an $(s\times s)$ order mathematical matrix is accomplished by the voltage amplifier μ_k, Fig. 5, of voltage amplification μ with its input connected to the *k*th row of the synthesis network, and its output of bipolarity connected to the *k*th column, together with suitable self and mutual admittances. The nondiagonal entries of the mathematical matrix are represented by the dinode admittances $y_{1k\pm}$, $y_{2k\pm}$, and so on, and the diagonal term or k,kth entry is represented by the dinode admittances y_{kk_m}.

An ensemble of these synthesis networks with the addition of suitable prescribed current sources j_k corresponding to the prescribed known vector in the mathematical matrix, as indicated in Fig. 6, then completes the synthesis of the given mathematical matrix equation.

The generalized electronic network admittance equation of the ensemble of basic synthesis columns, such as those corresponding to an $(s\times s)$ order mathematical matrix, reads:

$$([1_y]+[2_y]+[3_y]+[4_y])\cdot\begin{bmatrix}v_1\\v_2\\\cdot\\v_k\\\cdot\\v_r\\v_s\end{bmatrix}=\begin{bmatrix}j_1'\\j_2'\\\cdot\\j_k'\\\cdot\\j_r'\\j_s'\end{bmatrix}\qquad(20)$$

in which the ideal synthesis matrix is:

$$[1_y]=\begin{bmatrix}(y_{11+}-y_{11-})\mu_1\ldots(y_{1k+}-y_{1k-})\mu_k\ldots(y_{1s+}-y_{1s-})\mu_s\\(y_{21+}-y_{21-})\mu_1\ldots(y_{2k+}-y_{2k-})\mu_k\ldots(y_{2s+}-y_{2s-})\mu_s\\\vdots\qquad\ldots\qquad\cdot\qquad\ldots\qquad\cdot\\(y_{k1+}-y_{kl-})\mu_1\ldots(y_{kk+}-y_{kk-})\mu_k\ldots(y_{ks+}-y_{ks-})\mu_s\\\vdots\qquad\ldots\qquad\cdot\qquad\ldots\qquad\cdot\\(y_{r1+}-y_{r1-})\mu_1\ldots(y_{rk+}-y_{rk-})\mu_k\ldots(y_{rs+}-y_{rs-})\mu_s\\(y_{s1+}-y_{s1-})\mu_1\ldots(y_{sk+}-y_{sk-})\mu_k\ldots(y_{ss+}-y_{ss-})\mu_s\end{bmatrix}\qquad(21)$$

and the network error matrix reads:

$$[2_y]=\begin{bmatrix}\sum_{i=1}^{s}(y_{1i+}+y_{1i-})\ldots & 0 & \ldots & 0\\0 & \ldots & 0\quad\ldots & 0\\\vdots & \ldots & \vdots\quad\ldots & \vdots\\0 & \ldots\sum_{i=1}^{s}(y_{ki+}+y_{ki-}) & & 0\\\vdots & \ldots & \vdots\quad\ldots & \vdots\\0 & \ldots & 0\quad\ldots & 0\\0 & \ldots & 0 & \sum_{i=1}^{s}(y_{si+}+y_{si-})\end{bmatrix}\qquad(22)$$

The equivalent source error matrix $[{}^3_y]$ and the parasitic error matrix $[{}^4_y]$ may or may not be significant, depending upon the magnitudes of the particular problem, but can be made as small as desired. Finally, the right-hand column vector of j':s in equation 20 represents the current-source equivalents actually constructed from intrinsic voltage sources and series impedances.

Applying to equation 20 the transformation:

$$[v]=C\cdot(-w)\qquad(23)$$

where

$$C=\begin{bmatrix}\frac{1}{\mu_1} & 0\ldots.0\ldots.0 & 0\\0 & \frac{1}{\mu_2}\ldots0\ldots.0 & 0\\\vdots & \cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot & \cdot\\0 & 0\ldots.\frac{1}{\mu_k}\ldots0 & 0\\\vdots & \cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot & \cdot\\0 & 0\ldots.0\ldots.\frac{1}{\mu_r} & 0\\0 & 0\ldots.0\ldots.0 & \frac{1}{\mu_s}\end{bmatrix}\qquad(24)$$

and the vectors $[v]$ and $[-w]$ are defined in Fig. 5, results in the expression:

$$([0_y]+[2_y]C+[3_y]C+[4_y]C)\cdot\begin{bmatrix}-w_1\\-w_2\\\cdot\\-w_k\\\cdot\\-w_r\\-w_s\end{bmatrix}=\begin{bmatrix}j_1'\\j_2'\\\cdot\\j_k'\\\cdot\\j_r'\\j_s'\end{bmatrix}\qquad(25)$$

It should be noted that the $-w$:s represent the voltage outputs of the amplifiers from which readings are obtained of the solutions to the mathematical problem.

In equation 25, the term

$$[0_y]=[1_y]C\qquad(26)$$

which can now be made to correspond exactly to the mathematical problem matrix, reads:

$$[0_y]=\begin{bmatrix}(y_{11+}-y_{11-})\ldots(y_{1k+}-y_{1k-})\ldots(y_{1s+}-y_{1s-})\\(y_{21+}-y_{21-})\ldots(y_{2k+}-y_{2k-})\ldots(y_{2s+}-y_{2s-})\\\vdots\qquad\ldots\qquad\cdot\qquad\ldots\qquad\cdot\\(y_{k1+}-y_{k1-})\ldots(y_{kk+}-y_{kk-})\ldots(y_{ks+}-y_{ks-})\\\vdots\qquad\ldots\qquad\cdot\qquad\ldots\qquad\cdot\\(y_{r1+}-y_{r1-})\ldots(y_{rk+}-y_{rk-})\ldots(y_{rs+}-y_{rs-})\\(y_{s1+}-y_{s1-})\ldots(y_{sk+}-y_{sk-})\ldots(y_{ss+}-y_{ss-})\end{bmatrix}\qquad(27)$$

and no longer exhibits the amplifier voltage gains, μ:s.

Furthermore, the entries in the matrices $[{}^2_y]\cdot C$, $[{}^3_y]\cdot C$, and $[{}^4_y]\cdot C$ may be made as small as desired by making the respective μ:s sufficiently large.

The stability of the electronic synthesis network depends upon the frequency response of the amplifier and is discussed in reference 3. Briefly, the network has an equation of state which is always of differential-equation type, due to the amplifier characteristics and stray admittances, even for the synthesis of a purely algebraic mathematical problem. The stability of the synthesis network thus depends not only upon the mathematical problem to be solved, but is in fact heavily influenced by the amplifier characteristics and stray admittances as mentioned. Intentionally placing the capacitances on the principal diagonal terms, as indicated in the paper, tends to mask these effects and thus to yield a stable synthesis network, provided the problem to be solved is itself mathematically stable, of course.

Finally, therefore, to the degree of approximation desired, the electronic synthesis is described by the matrix admittance equation:

$$[0_y].[-w]=[j']\qquad(28)$$

and is now in the desired form, and homeomorphic, entry per entry, to a purely mathematical matrix equation to be solved.

References

1. Matrices in Analogue Mathematical Machines, Pierre M. Honnell, Robert E. Horn. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 260, no. 3, Sept. 1955, pp. 193–207.

2. Matrices in Electronic Differential Analyzers, Pierre M. Honnell, Robert E. Horn. *Proceedings*, International Analogy Computation Meeting, Brussels, Belgium, 1956, pp. 217–21.

3. Analogue Computer Synthesis and Error Matrices, Pierre M. Honnell, Robert E. Horn. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, Mar. 1956, pp. 26–32.

4. Synthesis of Vector Networks, R. E. Horn, V. G. Fauque. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., vol. PGEC-6, no. 4, Dec. 1957, pp. 261–65.

5. Matrix Programming of Electronic Analog Computers, R. E. Horn, P. M. Honnell. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, Sept. 1958, pp. 420–28.

6. An Electronic Simultaneous Equation Solver, E. A. Goldberg, G. W. Brown. *Journal of Applied Physics*, New York, N. Y., vol. 19, Apr. 1948, pp. 339–45.

7. Stabilization of Simultaneous Equation Solvers, G. A. Korn. *Proceedings*, Institute of Radio Engineers, vol. 37, Sept. 1949, pp. 1000–02.

8. An Automatic Simultaneous Equation Computer and Its Use in Solving Secular Equations, W. A. Adcock. *Review of Scientific Instruments*, New York, N. Y., vol. 19, Mar. 1948, pp. 181–87.

9. Treatment of Linear Algebraic Systems on an Electronic Differential Analyzer, F. L. Gephart. *REAC Information Bulletin no. 3*, Reeves Instrument Corporation, New York, N. Y.

Effects of Ultrahigh Temperature on Magnetic Properties of Core Materials

MICHAEL PASNAK
NONMEMBER AIEE

RICHARD LUNDSTEN
NONMEMBER AIEE

Synopsis: An experimental study has been made of the effects of temperature on the magnetic properties of the following ferromagnetic alloys: Orthonol, *4-79* Mo Permalloy, AEM *4750*, *L* and *Z* Silectron, Transformer *A*, Audio Transformer *A*, *11.7* Alfenol, *15.5* Alfenol, *3* Mo Thermenol, *7-70* Perminvar, and Supermendur. Measurements were made of the d-c and 60 cps (cycles per second) magnetic properties in the temperature range of 24–800 C (degrees centigrade). The materials were also evaluated after several temperature cycles between 24–500 C. The results indicate that, in general, high temperature decreased the maximum induction and the residual induction from their 24 C values. For all materials tested, with the exception of Supermendur and *11.7* Alfenol, the coercive force decreased with increase in temperature. The maximum permeability increased with increasing temperature until the Curie temperature was approached, then it started to decrease. The initial permeability also increased with increasing temperature, but unlike the maximum permeability, it generally started to decrease at a greater rate and at a lower temperature. Grain-oriented materials were more affected by temperature cycling than were unoriented materials. The effects of temperature cycling on the magnetic properties generally appeared as an increase in the coercive force, a decrease in the maximum permeability, a decrease in the residual flux density, and a decrease in the squareness ratio.

A STUDY of the effects of large temperature variations on the properties of magnetic core materials was made to facilitate the development of military components for operation under extreme environmental conditions.[1,2] The materials chosen are used in transformers, magnetic amplifiers, relays, and servo components.

This paper describes the effect of ambient temperature variations on the d-c and 60-cps magnetic properties of 30 specimens, representing 12 different magnetic alloys. These alloys are Transformer *A*, Audio Transformer *A*, *L*-Silectron, *Z*-Silectron, Orthonol, AEM *4750*, *4-79* Mo Permalloy, *11.7* Alfenol, *15.5* Alfenol, *3* Mo Thermenol, Supermendur, and *7-70* Perminvar. The data concerning core composition and geometry are given in Table I.

Experimental Procedure

The cores, all toroidal, were encased in uncovered type *304* stainless steel core boxes. Type *304* stainless steel was used because it can withstand the high temperatures reached in this experiment and is "nonmagnetic."[3] To eliminate any possibility of straining the specimen as a result of tightness in the core box, the core boxes were made oversize to allow for changes in core and box dimensions with temperature.

The toroidal laminations were coated with magnesium oxide and further insulated from each other by 0.004-inch Fiberglas laminations, while the convolutions of the spiral-wound tape cores were insulated only with magnesium oxide. The cores, in their stainless-steel core boxes, were wound with Silotex insulated wire supplied by Anaconda Wire and Cable Company.

The temperature study consisted of two phases. In the first phase, all the materials were tested at temperatures up to 500 C in 100 C steps or until the Curie point of the material was exceeded, if the Curie point was below 500 C. In order to insure temperature stability, the materials were allowed to remain at each temperature for a period of one hour before being tested. Inasmuch as it required one hour for testing, the total time at any given temperature was 2 hours. After being tested at 500 C or at the Curie temperature, whichever was lower, the material was allowed to furnace cool back to room temperature; (see Fig. 1, curve *A*). After the room temperature retest, the material temperature was raised directly to 500 C, or to its Curie temperature, and allowed to stabilize for one hour before retesting at this temperature. This cyclic procedure and testing was repeated three times. The core was then furnace cooled from 500 C or the Curie temperature at a faster cooling rate by removing the furnace cover; see Fig. 1, curve *B*.

The second phase of the temperature study involved the testing of only those materials whose Curie temperatures were higher than 500 C. New specimens of these materials were tested at 24 C and taken directly to 500 C where they were allowed to temperature stabilize for one hour before being tested. Then the temperature was raised from 500 C, in steps of 100 C, until the Curie temperature was exceeded for each material, and the materials were tested at each step. The materials were then allowed to furnace cool along curve *A* in Fig. 1, after which they were tested at 24 C.

Standard d-c test methods were used to obtain the static magnetic properties. The 60-cps dynamic loops were obtained by using a new instrument called the Ferrotracer.[4,5] Unless specified, the 60-cps dynamic loops shown in this report were obtained with the test core driven by a sine-wave "current source."

Table I. Core Materials Used in Temperature Study

Specimen	Composition, %	Core Construction	Inner Diameter, Inches	Outer Diameter, Inches	Thickness, Inches
Transformer *A*	4.7 Si+95.3 Fe	Toroidal laminations	$1^{5}/_{8}$	2	0.015
Audio transformer *A*	3.6 Si+96.4 Fe	Toroidal laminations	$1^{5}/_{8}$	2	0.014
L Silectron	3 Si+97 Fe	Spiral tape	$1^{5}/_{8}$	2	0.004
Z Silectron	3 Si+97 Fe	Spiral tape	$1^{5}/_{8}$	2	0.004
Orthonol	50 Ni+50 Fe	Spiral tape	$1^{5}/_{8}$	2	0.002
Orthonol	50 Ni+50 Fe	Square laminations	1.1875*	1.750†	0.002
AEM *4750*	47–50 Ni+47–50 Fe	Toroidal laminations	$1^{5}/_{8}$	2	0.014
4-79 Mo Permalloy	4 Mo+79 Ni+17 Fe	Toroidal laminations	$1^{5}/_{8}$	2	0.014
11.7 Alfenol	11.7 Al+88.3 Fe	Toroidal laminations	$1^{5}/_{8}$	2	0.014
15.5 Alfenol	15.5 Al+84.5 Fe	Toroidal laminations	$1^{5}/_{8}$	2	0.014
3 Mo Thermenol	3 Mo+16 Al+81 Fe	Toroidal laminations	$1^{5}/_{8}$	2	0.014
Supermendur	2 V+49 Co+49 Fe	Spiral tape	$1^{5}/_{8}$	2	0.002
7-70 Perminvar	7 Co+70 Ni+23 Fe	Toroidal laminations	$1^{5}/_{8}$	2	0.014

* Side of inner square.

† Side of outer square.

Paper 59-1117, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted August 19, 1958; made available for printing August 21, 1959.

MICHAEL PASNAK and RICHARD LUNDSTEN are with the United States Naval Ordnance Laboratory, Silver Spring, Md.

The authors gratefully acknowledge the assistance of J. H. Powers for his aid in the performance of the d-c testing and of D. I. Gordon for his helpful suggestions and criticisms in the preparation of the manuscript.

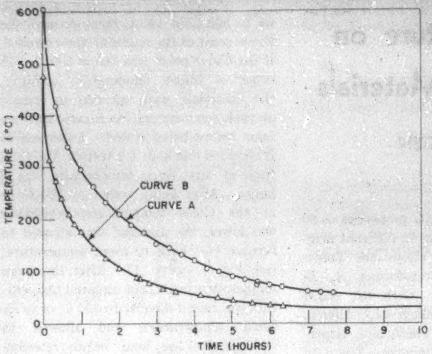


Fig. 1. Furnace cooling characteristics

O With cover on
△ With cover off

Results and Discussion

The influence of temperature, temperature cycling, and cooling rate is illustrated in Figs. 2–16.

SILICON-IRON

Of the four silicon-iron materials tested, the grain-oriented alloys *L* and *Z* Silectron, were more affected than the unoriented silicon-iron alloys, Audio Transformer *A* and Transformer *A*. The effects of temperature (up to 500 C) upon the d-c induction curves of the foregoing materials are shown in Fig. 2, where *L*-Silectron is grain oriented and Audio Transfomer *A* is unoriented. From this figure, one may well conclude that even though the grain-oriented material displayed greater over-all changes than the unoriented material, its properties are still superior to those of the unoriented material during the course of one temperature cycle. However, this superiority decreases with each temperature cycle as indicated by the behavior of the d-c induction curves of *Z*-Silectron and Transformer *A* shown in Fig. 3. In this figure, it is seen that temperature cycling from 24 C to 500 C and back to 24 C has little or no effect on the unoriented silicon-iron core. Whereas the same temperature cycling performed on the grain-oriented silicon-iron core resulted in a continuous degradation of its induction curves until they are comparable to those of the unoriented silicon-iron core.

The magnetic properites of the grain-oriented silicon-irons deteriorated with each temperature cycle in the 24–500 C range, indicating that the temperature was high enough to cause some degree of disorientation with each cycle. That disorientation had begun to take place is further evidenced when comparing the dynamic loops above 500 C of the grain-oriented material, Fig. 4, with those of the unoriented material, Fig. 5. By direct comparison of these two figures, it is seen that the dynamic loops of these two materials are very similar in this temperature range, and that disorientation of the grain-oriented material is almost complete. The grain-oriented material still displayed superior magnetic properties upon returning to 24 C from 800 C since it may not have been at the high temperature for sufficient time for permanent changes to occur.

Another factor very probably contributing to the deterioration of the magnetic properties of the silicon-iron alloys was the severe oxidation which occurred along the edges of the tape core and along the edges of the toroidal laminations. This oxidation of material, appearing as a heavy caked coating of rust, would decrease the effective area of the core material and produce a decrease in the maximum flux. This decrease in the effective area was not taken into account when calculating the maximum and residual flux densities.

NICKEL-IRON

The effect of temperature upon the 60-cps dynamic loops of the grain-oriented nickel-iron, Orthonol, shown in Fig. 6, indicated only a very small loss of rectangularity with increasing temperature.

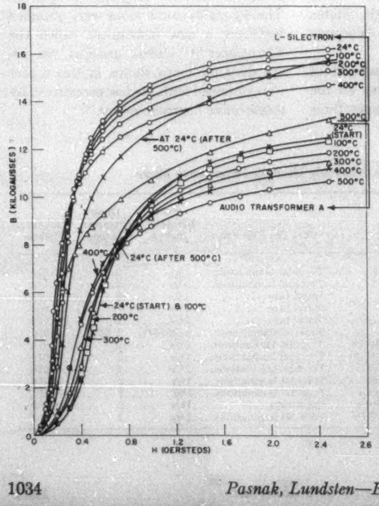
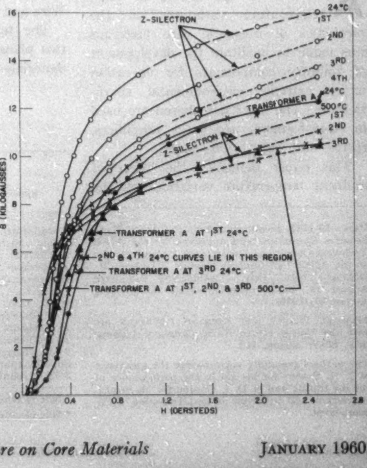


Fig. 2 (left). D-c induction curves of L Silectron and audio transformer A in the temperature range of 24–500 C

Fig. 3 (right). Effects of temperature cycling (24–500 C) on the d-c induction curves of Z Silectron and transformer A



Vol.
Pt. 1
195

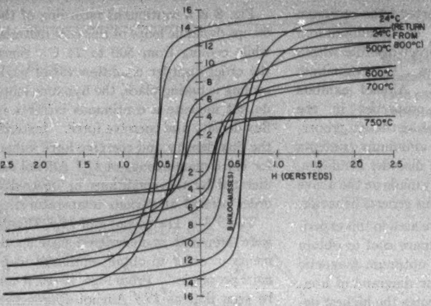


Fig. 4. 60-cps dynamic loops of L-Silectron in the temperature range of 24–750 C

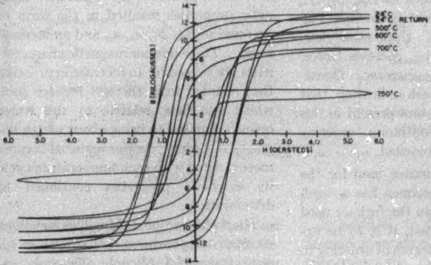


Fig. 5. 60-cps dynamic loops of transformer A in the temperature range of 24–750 C

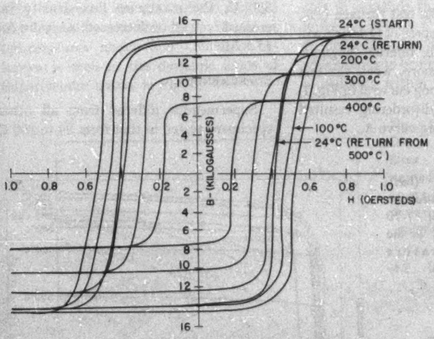


Fig. 6. 60-cps dynamic loops of Orthonol (tape) in the temperature range of 24–400 C

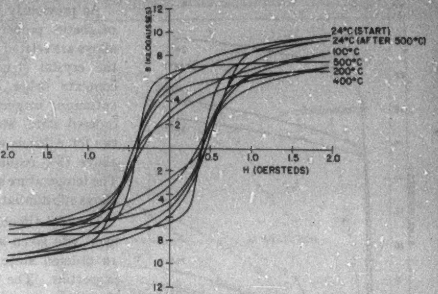


Fig. 7. 60-cps dynamic loops of 11.7 Alfenol in the temperature range of 24–500 C

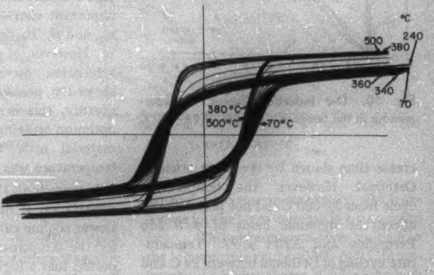


Fig. 8. Continuous 60-cps dynamic loops of 11.7 Alfenol while cooling from 500–24 C

At 380 C Bm and Br stopped increasing and started decreasing

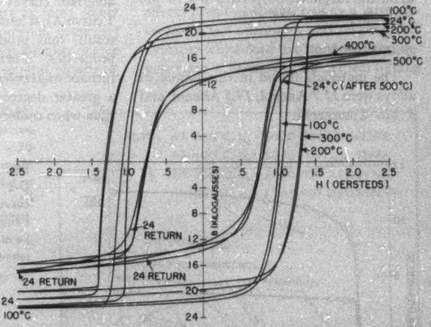


Fig. 9. 60-cps dynamic loops of Supermendur in the temperature range of 24–500 C

Upon returning to 24 C from 500 C a new dynamic loop was obtained which differs from the initial 24 C loop in that it has both a smaller coercive force and maximum flux density for the same field intensity. The decrease in a-c coercive force is probably an indication that some degree of disorder in the crystal structure had occurred. The d-c induction curves indicated an increase in d-c coercive force implying that some disorientation had taken place. Disorientation increases the d-c coercive force and decreases the maximum flux density.

The 60-cps dynamic loops of the unoriented nickel-iron materials (AEM *4750* and *4-79* Mo Permalloy) indicated a greater deterioration with temperature in-

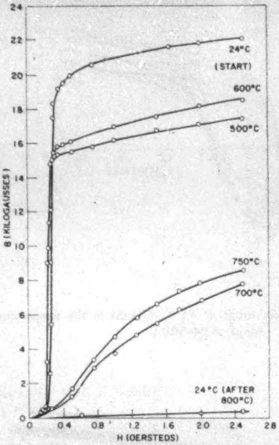


Fig. 10. D-c induction curves of Supermendur in the temperature range of 24–750 C

crease than shown by the grain oriented Orthonol. However, the temperature cycle from 24–500 C did not permanently affect the dynamic loops of *4-79* Mo Permalloy and AEM *4750*. Temperature cycling of Orthonol between 24 C and 500 C had a very pronounced effect upon its maximum permeability and coercive force. Whereas, the same temperature cycling had little or no effect upon the magnetic properties of *4-79* Mo Permalloy and AEM *4750*.

ALUMINUM-IRON

The materials tested in this family of alloys were *11.7* Alfenol, *15.5* Alfenol, and *3* Mo Thermenol.

As previously reported,[6] the optimum magnetic properties of aluminum-iron alloys are related to order-disorder within the crystal structures. For aluminum contents below 13%, Alfenol exhibits optimum magnetic properties in the ordered state, whereas optimum properties for Alfenols with aluminum contents above 13% require a disorder condition. The temperature study made on the above alloys substantiates this general behavior.

The *11.7* Alfenol core used in this experiment was given a furnace cool to obtain an ordered state for optimum magnetic properties. The phase diagram[7] of iron-aluminum alloys indicates the highest degree of ordering of *11.7* Alfenol occurs at approximately 550 C. As the temperature was increased from 24 C to 400 C such important characteristics as B_m, B_r, H_c, u_m, and B_r/B_m decreased as shown in Fig. 7. However, at 500 C, these same characteristics, except the coercive force, showed a pronounced increase. Consequently, this is a good indication that optimum ordering was not present in this material until the optimum ordering temperature was approached.

Inasmuch as the furnace used for the preparation of this specimen has a much slower cooling rate than the furnace used for the temperature test, it is to be expected that a higher degree of ordering in the specimen would be obtained in the former furnace. Therefore, it can be expected that upon returning to 24 C at a faster rate, as given by the cooling characteristic curves of the test furnace (Fig. 1 curve A), a lesser degree of ordering will result, thus producing inferior magnetic properties. Furthermore, when the test furnace was cooled along curve B in Fig. 1, a greater degree of disordering resulted than when cooled along curve A.

Fig. 8 is a continuous recording of the 60-cps dynamic loops of this core material while cooling from 500 to 24 C. Since the order-disorder condition varies while cooling is taking place, the dynamic loops do not indicate a continuous increase in flux density and coercive force. Instead, the flux density and coercive force values for the dynamic loops of *11.7* Alfenol are dictated by the conditions or order-disorder present at various temperatures.

The *3* Mo Thermenol and *15.5* Alfenol were prepared as disordered state specimens in order to obtain optimum magnetic properties. From Figs. 12–16, it can be seen that as *15.5* Alfenol approached higher temperatures, a greater degree of ordering took place and the magnetic properties deteriorated. When this material was returned to 24 C inferior magnetic properties resulted in the form of decrease in B_m, B_r, u_o, u_m and an increase in H_c. These inferior magnetic characteristics are caused by an increase in ordering that resulted from the test furnace cool, which was slow, relative to the water quench required to bring about a high degree of disorder. As this material became more ordered with each temperature cycle its magnetic properties continued to deteriorate.

The behavior of *3* Mo Thermenol at elevated temperatures was similar in all respects to *15.5* Alfenol. The only exceptions are that the Curie temperature of *3* Mo Thermenol was approximately 300 C and that upon returning to 24 C from 300 C, the maximum flux density increased instead of decreased, as it did for *15.5* Alfenol.

SUPERMENDUR

Supermendur differed from all other specimens tested in that from 24 to 300 C

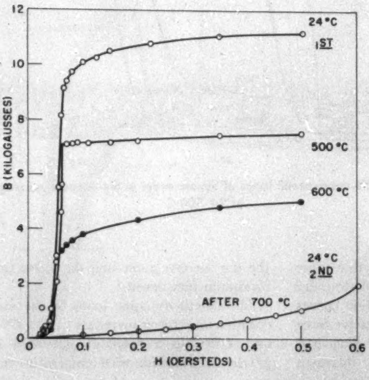
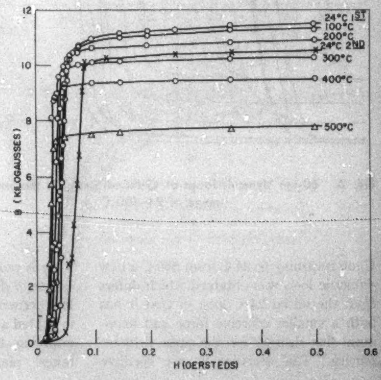


Fig. 11 (left). D-c induction curves of 7-70 Perminvar in the temperature range of 24–600 C

Fig. 12 (right). D-c induction curves of 7-70 Perminvar in the temperature range of 24–500 C



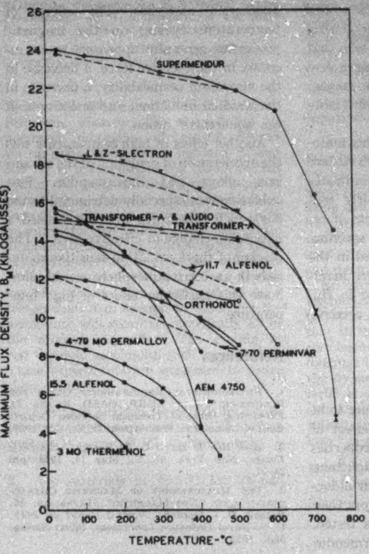


Fig. 13 (left). D-c maximum flux density at 30 oersteds versus temperature curves of all materials tested

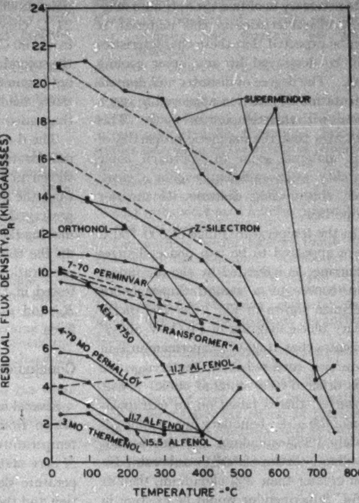


Fig. 14 (right). D-c residual flux density at 30 oersteds versus temperature curves of all materials tested

there was approximately a 10% decrease in the d-c coercive force and a 30% increase in the a-c coercive force. From 300 C to 500 C, the d-c coercive force increased approximately 10%, while the a-c coercive force decreased approximately 42%. Further, the d-c and a-c residual flux densitites showed sharp drops at temperatures above 300 C. This behavior is seen in the dynamic loops of Supermendur shown in Fig. 9.

In addition to being dependent upon domain orientation for optimum magnetic properties, the magnetic properties of Supermendur are also critically dependent upon the cooling rate to which it is subjected. In view of these facts, it is reasonable to conclude that in the temperature range of 24–300 C there is most likely an order-disorder phenomenon occurring which tends to bring about a higher degree of ordering within the crystal structure, causing a decrease in resistivity, which accounts for the large increase in a-c coercive force. This phenomenon reverses itself in the temperature range of 300–500 C bringing about a higher degree of disorder which increases the resistivity, thus decreasing the a-c coercive force. Apparently, the effect of temperature upon domain orientation is not felt until 400 C where its effects manifest themselves as a large decrease in the maximum and residual flux densities which lead to a very poor squareness ratio. Because of the critical cooling

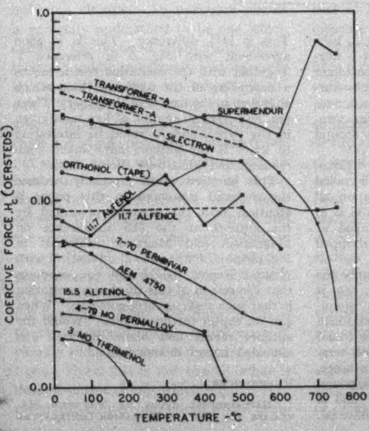


Fig. 15 (left). D-c coercive force at 30 oersteds versus temperature curves of all materials tested

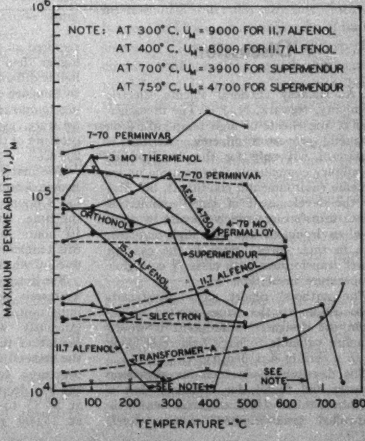


Fig. 16 (right). D-c maximum permeability versus temperature curves for all materials tested rate necessary to obtain the optimum magnetic characteristics in this material, it can be expected that these characteristics will be destroyed by any other cooling rate. The degree of disorder and domain orientation destruction appeared to reach a limit with each temperature cycle. This indicates that the magnetic properties of this material seem to approach some stability after sufficient cycling. However, this cycling destroys its superior properties

In the temperature range of 500–750 C there appeared to be two phase changes occurring, as indicated by the d-c induction curves of a second Supermendur specimen shown in Fig 10. The quasi-binary phase diagram of Fe-Co-V alloys[7] indicates that at 500 C, Supermendur is in an $\alpha'+\gamma$ (ordered-body centered cubic plus ordered-face centered cubic) phase. However, this is based on an extrapolation of the $\alpha'+\gamma$ boundary line to approximately 1.8% vanadium. If this boundary line is extrapolated to any composition greater than 2% vanadium, then at 500 C Supermendur would be existing in the α' phase, and at 600 C in the $\alpha'+\gamma$ phase, at 700 C in the α' phase, and at 750 C in the α phase. The behavior of the induction curves of Fig. 10 indicates that the $\alpha'+\gamma$ boundary in the quasi-binary phase diagram in reference should be extrapolated to a vanadium content of greater than 2% in order to account for their behavior.

7-70 PERMINVAR

Like Supermendur, 7-70 Perminvar depends upon cooling rate for the desired magnetic properties, especially when cooled from temperatures above its Curie point as shown in Fig. 11 where the second 24 C curve was obtained after cooling from 700 C. In this figure, it is seen that the second 24 C d-c induction curve had not begun to saturate for the same magnetizing fields, which previously were more than adequate for saturation.

The d-c induction curves of this material for temperatures up to 500 C are shown in Fig. 12. This figure indicates that the rectangularity was fairly well preserved. In Fig. 11, the effects of removing the magnetic anneal and speeding up the cooling rate over the used in the preparation of this specimen were manifested in the form of a decrease in B_m, B_r, and u_m and an increase in coercive force.

Conclusions

Several important conclusions are to be drawn from this study of the effect of temperature on the magnetic properties of core materials. In general, high temperature decreased the maximum induction and the residual induction from their 24 C values. For all materials tested, with the exception of Supermendur and *11.7* Alfenol, the coercive force decreased with increase in temperature. The maximum permeability increased with increasing temperature until the Curie temperature was approached, then it started to decrease. The initial permeability also increased with increasing temperature, but unlike the maximum permeability, it generally started to decrease at a greater rate and at a lower temperature.

Grain-oriented materials were more affected by temperature cycling than were unoriented materials. The effects of temperature cycling on the magnetic properties generally appeared as an increase in the coercive force, a decrease in the maximum permeability, a decrease in the residual induction, and a decrease in the squareness ration.

Another effect of high temperature was the acceleration of oxidation of the silicon-iron alloys and Supermendur. This oxidation was especially detrimetal to the 0.004-inch tape cores where it caused a noticeable change in effective area. The placing of these oxidation sensitive materials in an inert atmosphere would eliminate this particular effect of high temperature.

References

1. ENVIRONMENTAL REQUIREMENTS GUIDE FOR ELECTRONIC PARTS, AGEP, OASD, R AND D. *PB 131423*, Office of Technical Services, Department of Commerce, Washington, D. C., Oct. 1957.

2. HOW HOT IS HOT? **G. H. Rostky.** *Electronic Design*, New York, N. Y., May 14, 1958, pp. 24–27.

3. THE MEASUREMENT OF MAGNETIC CHARACTERISTICS OF "NON-MAGNETIC" MATERIALS, **M. Pasnak, D. I. Gordon.** *NavOrd Report no. 2415*, U. S. Naval Ordnance Laboratory, Silver Spring, Md., 1952.

4. THE FERROTRACER—A SIMPLE DYNAMIC B-H LOOP TRACER WITH SILICON-DIODE CHOPPER CIRCUITS, **W. A. Geyger.** *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 12, Oct. 1956, pp. 83–97.

5. MAGNETIC-AMPLIFIER CIRCUITS (book), **W. A. Geyger.** McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1957, chap. 20, pp. 350–69.

6. THE FABRICATION AND PROPERTIES OF ALFENOL ALLOYS (AL-FE) CONTAINING 10 TO 17% ALUMINUM, **J. F. Nachman, W. J. Buehler.** *NavOrd Report no. 4130*, U. S. Naval Ordnance Laboratory, 1955.

7. FERROMAGNETISM (book), **R. M. Bozorth.** D. Van Nostrand Company, Inc., Princeton, N. J., 1951, fig. 7-1, p. 211.

8. *Ibid.*, figs. 6–12, p. 201.

◆

Discussion

A. A. Halacsy (Federal Pacific Electric Company, Newark, N. J.): The investigation of the effects of high temperature on magnetic properties of core materials is important not only for the development of military components for operation under extreme environmental conditions, but also for the development of quite commercial power transformers which operate in quite usual environment but with high temperature in their cores. Such is the case in power transformers with class *H* or higher temperature insulation.

The authors' report shows hysteresis loops all up to the same oersted value. From the designers' standpoint it would be more valuable to have hysteresis loops up to a certain flux density.

The discusser made such investigations recently and found, for instance, that a 0.012-inch-thick *M6X* material which is cold-rolled grain-oriented silicone steel, required an exciting current at 233 C, 10% higher than at 20 C. This is not a negligible difference, and it warns the customer to compare exciting currents at the operating temperature and not at the cold temperature, especially if the operating temperature of the steel core is high, say around 200 C.

The material was heat cycled five times. It was heated for 16 hours and air-cooled for 8 hours with 60-cps excitation on all the time. It reached the steady 233 C in 10 hours during the heating, stayed at this temperature for 6 hours, then dropped to 65 C during the next 8 hours.

For a transformer designer not only the hysteresis loop but the loss curve, too, is important. We made loss-tests simultaneously with the already mentioned excitation tests and, of course, we found the losses decreasing with increased temperature. In a typical case, for instance, the loss decreased from 0.562 watts/pound at 20 C to 0.500 watt/pound at 233 C at 15,000 gauss in an *M6X* material. This is 11% difference. This is in good agreement with Fig. 5 of reference 1. Together with the excitation test it means a narrowing of the hysteresis loop which shows up in the figures of the paper. There are, however, no results, of loss-tests given in the paper, and it would be interesting to know if the authors have taken such readings and what were their results.

This question arises especially because the material presented in this paper is identical with the material presented by the authors at the 4th Conference on Magnetism and Magnetic Materials in Philadelphia, Pa., and is also identical with Navord Report 6132,[2] and one supposes that the reason of this third representation is that such tests are still going on. When can we expect further results? Would the authors reveal how their material was annealed before it was tested?

REFERENCES

1. EXPERIMENTAL STUDY OF MAGNETIC MATERIALS FOR USE IN ULTRAHIGH-TEMPERATURE

ELECTRONIC TRANSFORMERS, H. B. Harms. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, May 1958, pp. 181–84.

2. *NavOrd Report no. 6132*, U. S. Naval Ordnance Laboratory, Silver Spring, Md., July 10, 1958.

Weston Morrill (General Electric Company, Pittsfield, Mass.): The degradation of magnetic properties of crystal-oriented silicon irons, when cycled between 24 and 500 C, is not caused by a disorientation within the grain-oriented material, but is caused by a chemical aging phenomenon. Interstitial elements, such as carbon, held in supersaturated solution, will precipitate at an accelerated rate with increasing aging temperatures, resulting in an attendant damage of magnetic characteristics. Carrying the material for short periods to a temperature high enough to redissolve the precipitate will recapture practically all of the original properties. The cycling at 800 C would accomplish such regeneration. These experiments are suggested to prove these points:

1. Chemical analysis for carbon by a low-pressure combustion method. If carbon content is over 0.001% chemical aging is possible.

2. Age materials at 150 C for 100-odd hours and establish the magnitude of magnetic degradation. It probably will be as great as any noted in the original work.

3. Recycle at the lower temperatures the material which has been regenerated by heating to 800 C, and note a duplicate degradation of properties.

Michael Pasnak and **Richard H. Lundsten:** The magnetizing fields used by the authors were such that the core materials were being driven well into saturation. Therefore, it was felt that the magnetizing field could be interpolated at any desired flux level from the submitted induction curves and dynamic loops.

No loss measurements were made with wattmeters or any of the acceptable bridge techniques. Here again, the authors felt that all the necessary data concerning total core losses were obtained in the form of the dynamic loops. Since the area enclosed by the dynamic loop of a magnetic core material is directly proportional to the total core losses, these losses can be obtained by substituting the area of the dynamic loop into the equation

$$W=\frac{KA}{4\pi}$$

where

K = to the product of gausses per unit ordinate and oersteds per unit abscissa

A = area of dynamic loop in cm^2

W = ergs/cm^3 cycle

This paper was not re-presented at the AIEE Fall General Meeting. It was merely submitted for discussion. Further studies pertaining to the effects of temperature on the magnetic properties of core materials have been discontinued.

Standard annealing temperatures and cooling rates were used for the preparation of the test specimens.

The authors would like to thank Dr. Halacsy for his discussion. We should also like to take this opportunity to suggest the use of the Ferrotracer as an invaluable instrument for obtaining magnetic data under dynamic conditions of magnetization.

Mr. Morrill's comments and suggestions are appreciated. The authors feel that contamination of the materials due to the air atmosphere used in their investigations may have masked the effects of precipitation of interstitial elements such as carbon. However, if time allows, in the future we will investigate Mr. Morrill's proposed suggestions.

A Comparison of Sequential and Iterative Circuits

E. J. McCLUSKEY, JR.
NONMEMBER AIEE

TRADITIONALLY, the major effort of switching theory has been devoted to finding techniques for synthesizing optimum circuits. An optimum circuit is generally understood to be one which exhibits the required relationship among its input and output signals, and also costs less to construct than any other such circuit. The requirements on the relation among input and output signals can include not only a specification of which output signals are to result from the various combinations or sequences of input signals, but also a restriction on the delay between the application of the input and the appearance of the corresponding output signal. The desired synthesis procedure is one which does not place any restrictions on the form of the resulting network. Thus, the final circuit is to be optimum for the entire class of circuits which satisfy the design specifications. Even this is not exactly what is desired since a circuit which requires very complicated interconnections would probably be undesirable because of difficulties involved in maintaining proper operation. To be completely accurate, the cost of the circuit should include some measure of the circuit complexity.

It is not too surprising that there has been a general lack of success in obtaining such synthesis procedures. Since the problem of designing optimum circuits is so difficult, attempts have been made to devise synthesis schemes which make use of greatly simplified measures of circuit cost. Thus a great deal of attention has been directed towards a procedure for designing 2-terminal contact networks in which the number of contacts and springs are minimized. The fact that extra springs on a relay may actually cost nothing is ignored as is also the effect of wiring complexity. In spite of the seeming simplicity of this problem it is still unsolved. Solutions do exist for certain other simplified problems; for example, that of obtaining a circuit constructed of AND-gates, OR-gates, and inverters which contains the minimum number of inverters,[1] and that of obtaining a 2-stage single-output diode-gate network containing the minimum number of gates.[2,3]

In practice, the design of switching circuits is usually done by making an initial assumption as to the general form of the final circuit. Specialized design techniques for circuits having the chosen form are then used to obtain the actual design. For example, special methods exist for designing circuits which have n-inputs and 2^n-outputs, only one of which is equal to 1 for each combination of input signals.[4] However, many of the special circuit configurations which are commonly used in practice have not been studied theoretically. When such circuits are to be designed it is necessary to resort to informal or trial and error techniques. This paper is discussion of one special form of circuit: a combinational circuit (a circuit which does not contain any storage or memory such as flip-flops, self-locking relays, etc; its present output is determined only by its present input).

Paper 59-1160, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted Feb. 6, 1959; made available for printing August 10, 1959.

E. J. McCLUSKEY, JR., formerly with Bell Telephone Laboratories, Inc., Whippany, N. J., is now with Princeton University, Princeton, N. J.

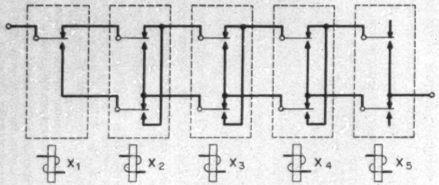

X1
X2
X3
X4
X5

Fig. 1 (left). A relay iterative network which is closed when an odd number of relays are operated

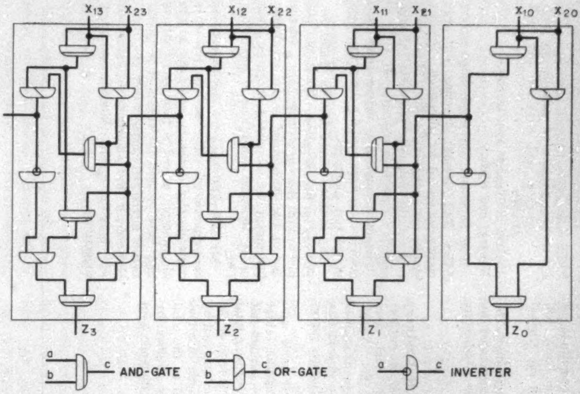

X13
X23
X12
X22
X11
X21
X10
X20
Z3
Z2
Z1
Z0
a
b
c
AND-GATE
OR-GATE
INVERTER

Fig. 2 (right). An iterative network for adding two binary numbers, X_1 and X_2

$$X_1 = 2^0X_{10} + \ldots + 2^3X_{13}$$
$$X_2 = 2^0X_{20} + \ldots + 2^3X_{23}$$
$$Z = X_1 + X_2$$

	s_j	ab 00	01	11	10	Z (OUTPUT)
A = B	1	①	3	①	2	0
A > B	2	②	②	②	②	0
A < B	3	③	③	③	③	1

(a) FLOW TABLE
(1 IS INITIAL STATE)

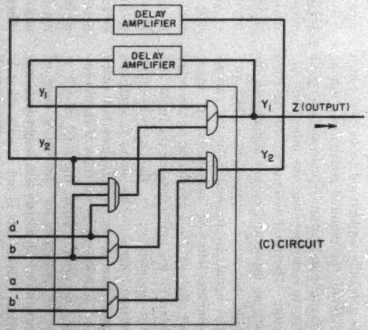
y_1	y_2	ab 00	01	11	10	Z (OUTPUT)
0	1	01	10	01	00	0
0	0	00	00	00	00	0
1	0	10	10	10	10	1

$Y_1 Y_2$

$Y_1 = y_1 + y_2 a'b;$

$Y_2 = y_2 (a+b')(a'+b);\ Z = Y_1$

(b) OUTPUT TABLE


DELAY AMPLIFIER
DELAY AMPLIFIER
y1
y2
Y1
Y2
Z (OUTPUT)
a'
b
a
b'

(C) CIRCUIT

Fig. 3 (left). A sequential circuit which has an output only if A<B

$$A = a_n2^n + \ldots + a_02^0$$
$$B = b_n2^n + \ldots + b_02_0$$

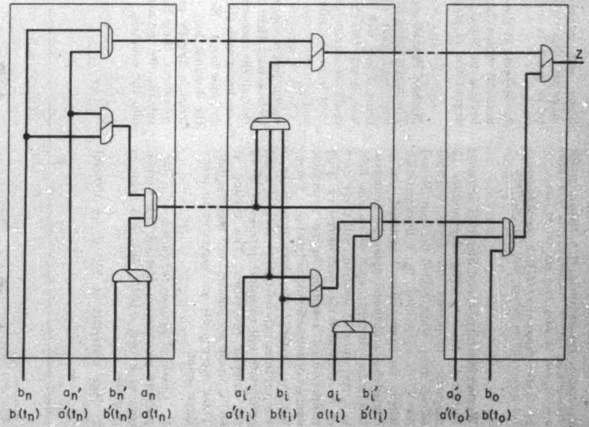
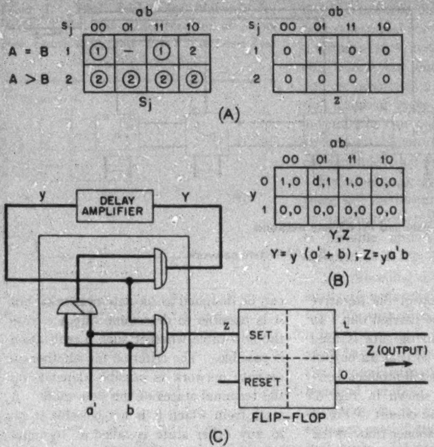

Z
b_n
a_n'
b_n'
a_n
$b(t_n)$
$a'(t_n)$
$b'(t_n)$
$a(t_n)$
a_i'
b_i
a_i
b_i'
$a'(t_i)$
$b(t_i)$
$a(t_i)$
$b'(t_i)$
a_o'
b_o
$a'(t_o)$
$b(t_o)$

Fig. 4 (right). An iterative circuit which performs the same function as the sequential circuit of Fig. 3



	s_j	00	01	11	10
A = B	1	①	–	①	2
A > B	2	②	②	②	②

s_j	00	01	11	10
1	0	1	0	0
2	0	0	0	0

y	00	01	11	10
0	1,0	d,1	1,0	0,0
1	0,0	0,0	0,0	0,0

$Y=y\ (a'+b)\ ;\ Z=ya'b$

Fig. 5 (left). Modification of the circuit of Fig. 3 to replace one delay amplifier with a flip-flop

A—Flow table
(1 is initial table)
B—Output table
C—Circuit

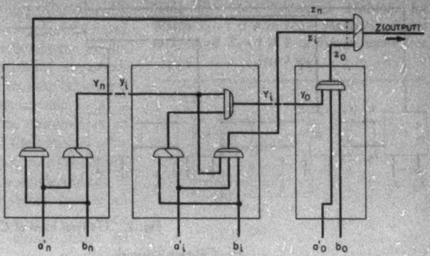


Fig. 6 (above). OR-gate network corresponding to Fig. 4

constructed as a series of identical "cells" or subnetworks. Such a circuit is commonly called an iterative circuit. Typical examples of iterative circuits are the parallel binary adder and the relay parity check circuit; see Figs. 1 and 2. A discussion of the reasons for using iterative networks has been given elsewhere.[5] Iterative circuits were first studied theoretically in the book by Keister, Ritchie, and Washburn.[6] In this book, an informal method is given for designing iterative contact networks. These networks are divided into two types: 1. symmetric circuits in which the outputs depend only on the number of relays operated, and 2. positional circuits in which the outputs depend on the relative positions of operated and unoperated relays. (A typical symmetric circuit is one which has output only if 2 or 4 of the inputs are equal to 1, and a typical positional circuit is one which has an output only if exactly two adjacent inputs are equal to 1.) The positional circuits include many important circuits having binary numbers as inputs.

Formal design procedures for iterative contact networks were proposed by Caldwell[7] and Huffman.[8] Huffman's method makes direct use of an analogy between iterative circuits and sequential circuits, and carries out the iterative network design by means of a flow table similar to that used in sequential circuit theory.[7,9] The method used by Caldwell is similar to Kleene's technique for sequential circuits,[10,11] but does not contain a formal procedure for going from a statement of the desired behavior to the actual circuit.

There are actually two general classes of iterative circuit: The first class consists of those circuits which have their outputs derived only from the final cell. These circuts will be called circuit-output iterative networks. Fig. 1 is an example of such a circuit. These are the types of circuits which have been studied theoretically by Caldwell and Huffman.

A large porportion of the iterative circuits which are actually used belong to the second class in which the circuit outputs are obtained from each of the individual cells. The circuit of Fig. 2 is an example of such a circuit. These circuits will be called cell-output iterative networks.

Design Procedure

The analogy between sequential and iterative circuits will be illustrated by means of an example. First a sequential circuit will be designed which compares two binary numbers, $A=a_na_{n-1}\ldots a_1a_0$ and $B=b_nb_{n-1}\ldots b_1b_0$; and has an output only if $A<B$. The first pulses to appear at the circuit inputs will correspond to a_n, a_n', b_n, b_n', the second pulses to a_{n-1}, a'_{n-1}, b_{n-1}, b'_{n-1}, etc. (a_n' represents the complement of a_n. A pulse will appear on the a lead if the corresponding bit is equal to 1 and on the a' lead if the corresponding bit is equal to 0.) The appropriate flow table and circuit are shown in Fig. 3.[13] The circuit diagram is incomplete since some additional circuitry would be necessary to insure coincidence of the input pulses and the y_1 and y_2 pulses from the delay amplifiers, and also to reset the circuit to the proper initial state. This circuitry has been omitted since it does not affect the logical structure of the circuit and simplifies the analogy with iterative circuits. It is assumed that the delay of the delay amplifiers is the same as the time interval between successive input pulses. All time in this circuit will be measured in terms of this interval between pulses. Thus, t_0 is the time when the first input pulses arrive, t_1 is the time when the second input pulses arrive, etc. Before t_0 the circuit is in its initial state, with $Y_1=y_1=0$ and $Y_1=y_2=1$. At t_0 the input pulses representing a_n and b_n arrive. If $a_n=0$ and $b_n=0$ or $a_n=1$ and $b_n=1$; the inputs to the delay amplifiers Y_1 and Y_2, remain unchanged. If $a_n=0$ and $b_n=1$, Y_1 becomes equal to 1 and Y_2 becomes equal to 0. If $a_n=1$ and $b_n=0$, Y_1 and Y_2 both become equal to 0. The values of y_1 and y_2 at t_1 will be equal to values of Y_1 and Y_2 at t_0. More simply $y_1(t_i+1)=Y_1(t_i)$ and $y_2(t_i+1)=Y_2\ (t_i)$ where $y_1(t_i+1)$ represents the value of y_1 at time t_i+1, etc. The values of $Y_1(t_i)$, $Y_2(t_i)$ and $Z(t_i)$ are determined by $a(t_i)$, $b(t_i)$, $y_1(t_i)$ and $y_2(t_i)$.

If all of the input bits: $a_n, a_{n-1}, \ldots, a_0$ and $b_n, b_{n-1}, \ldots b_0$, are available simultaneously (as d-c signals rather than pulses), it is possible to build an iterative network which performs the same function as this equential circuit. This is done by removing the delay amplifiers from the circuit of Fig. 3(C) and constructing n copies of the resulting circuit, Fig. 4. These copies of the sequential circuit form the cells of the iterative net-

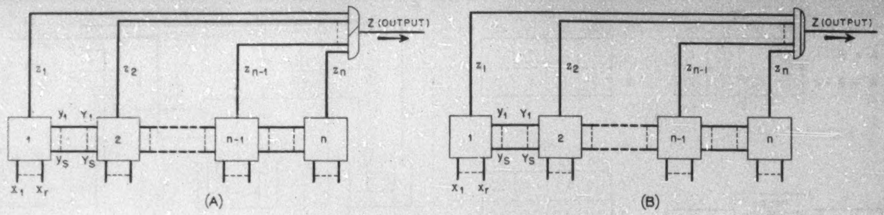


Fig. 7. **General form of OR-gate and AND-gate networks**

A—OR-gate network B—AND-gate network

work. Instead of being connected to the inputs of the delay amplifiers, Y_1 and Y_2 are now connected to the y_1 and y_2 inputs of the next cell of the iterative circuit. It is only necessary to derive an output from the final cell, since the circuit should have an output only if $Y_1=1$ and $Y_2=0$ for the final cell. This allows certain simplifications in the final cell since Y_1 and Y_2 need not be formed explicitly. It is very simple to verify that the iterative network will perform the same function as the sequential circuit. In both circuits: $y_i(t+1)=Y_i(t)$; and $Y_1(t)$ and $Y_2(t)$ are derived from $a(t)$, $b(t)$, $y_1(t)$, and $y_2(t)$ according to the same rules. This shows that an iterative network can be designed by the same techniques used to design sequential circuits.

Reduction of Delay in Forming an Output

In the circuit of Fig. 3, $Y_1=0$ and $Y_2=1$ as long as the inputs are either $a=0$ and $b=0$ or $a=1$ and $b=1$. As soon as some input for which $a=1$ and $b=0$ occurs, Y_2 becomes equal to 0, and Y_1 and Y_2 both remain equal to 0 regardless of the values of the subsequent inputs. Similarly, if the input $a=0$, $b=1$ occurs before $a=1$, $b=0$; Y_1 and Z become equal to 1 and Y_2 becomes equal to 0, and subsequent inputs have no effect on Y_1, Y_2, or Z. This makes possible an alternate realization of the sequential circuit in which a pulse occurs on a z-lead if the input $a=0$, $b=1$ occurs before the input $a=1$, $b=0$. This z pulse is used to set a flip-flop in the 1-condition and the circuit output is derived from the output of this flip-flop.[12] This circuit is shown in Fig. 5. The arrangements for synchronizing pulses and resetting the circuit to its initial condition have been omitted from the figure. A comparison of the economy of the circuits of Figs. 3 and 5 depends on the relative cost of delay amplifiers, flip-flop, and gate circuits.

A similar modification of the iterative circuit of Fig. 4 can be carried out. In the iterative circuit, an OR-gate is used instead of the flip-flop so that an output on the z-lead of any cell will produce a circuit output. This is shown in Fig. 6. In the modified iterative circuit of Fig. 6 the output will appear sooner than in the original circuit since the output can come directly from an intermediate cell rather than propagating through all of the remaining cells to the final cell. These can also be a saving in equipment in the modified circuit since $2n$ OR-gates each with two inputs are being replaced by a single OR-gate with n-inputs.

An iterative network in which the output is derived from an OR-gate which has one input from each cell of the network will be called an "OR-gate network." The general form of such a network is shown in Fig. 7(A). Not all iterative networks can be designed as OR-gate networks, but it is possible to determine directly from the flow-table whether such a realization is possible. The criterion for whether an OR-gate network is possible depends on the terminal states of the flow table. A state from which it is not possible to go to any other state is called a "terminal state." In the table of Fig. 3(A), states 2 and 3 are terminal states.

A circuit-output iterative network can be designed as an OR-gate network if and only if there is a terminal state for which the output is to equal 1. The design is obtained by deleting the terminal state from the flow table and generating a 1-output from a cell whenever the next state should be the terminal state which was removed. The table of Fig. 5(A) is obtained from the table of Fig. 3(A) by removing state 3 and generating a 1-output wherever the next-state entry of the

s_j \ ab	00	01	11	10
A=B 1	①	3	①	–
A<B 3	③	③	③	③

S_j

ab	00	01	11	10
	1	1	1	0
	1	1	1	1

z

(A)

y \ ab	00	01	11	10
	0,1	1,1	0,1	d,0
1	1,1	1,1	1,1	1,1

Y, z

$Y=y+a'b$, $z=y+a'+b$

(B)

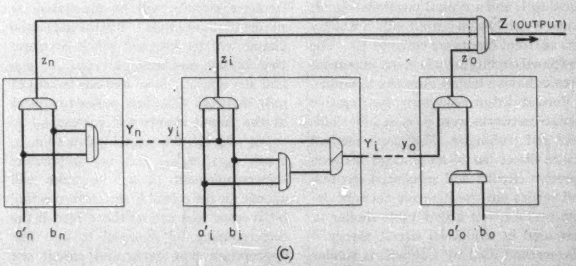


Fig. 8. **AND-gate network corresponding to Fig. 4**

A—Flow table (1 is initial table)
B—Output table
C—Circuit

Vol.
Pt. 1
195

Table I. State Tables for Network for Which $Z_1=1$ If an Odd Number of Inputs Are Equal to 1, and $Z_2=1$ If at Least Two Inputs Are Equal to 1

$s_j \backslash x_i$	0	1	Z_1	Z_2
1	1	2	0	0
2	2	3	1	0
3	3	4	0	1
4	4	3	1	1

S_j

(A)

$s_j \backslash x_i$	0	1	Z_1
1	1,0	2,0	0
2	2,0	3,1	1
3	3,d	4,d	0
4	4,d	3,d	1

S_j, z_2

(B)

$s_j \backslash x_i$	0	1	Z_1
1	1,0	2,0	0
2	2,0	1,1	1

S_j, z_2

(C)

original table was 3 (for $a=0$, $b=1$, and present state 1). In Fig. 5(A) a dash is used for the next state entry when the present state is 1 and $a=0$, $b=1$; since whenever this condition occurs, a 1-output will be generated and the next-state will be unimportant. Once an output has been generated, any further circuit action is unimportant since it cannot remove the output. In actually designing the circuit a state must be chosen to replace the dash, but this is done so as to minimize the equipment in the cell. The dash entry in the state table thus becomes a "don't care" entry in the output table; see Fig. 5(B).

It is also possible to design a circuit-output network in which the circuit output is derived from an AND-gate which has one input per cell; see Fig. 7(B). Such a network is called an "AND-gate iterative network." A circuit-output network can be designed as an AND-gate network if and only if there is a terminal state for which the circuit is to have a 0-output. The procedure for designing such a circuit as an AND-gate network is directly analogous to designing an OR-gate network. The terminal state for which there is to be a 0-output is deleted and a 1-output is generated from all cells for which the output state is not the terminal state being removed. For the table of Fig. 3(A) this requires the removal of state 2 and an output to be generated for all conditions except when the input state is 1 and the inputs are $a=1$, $b=0$. The resulting flow table and circuit are shown in Fig. 8. It is not true that there will always be both a terminal state with a 1-output and a terminal state with a 0-output. A circuit which has one terminal state can only be designed as an OR-gate network if the terminal state corresponds to a 1-output and can only be designed as an an AND-gate network if the terminal state corresponds to a 0-output from the circuit.

For a network with a single output, if there are no terminal states, neither an OR-gate network nor an AND-gate network is possible and the circuit must be designed as in Fig. 4.

The preceding discussion applies specifically to networks with a single output. For a network with more than one output, it may be possible to realize the network as an OR-gate or AND-gate network even if the original network does not contain any terminal states. For multiple-output networks it is necessary to consider terminal sets of states rather than terminal states. "A terminal set" of states or "terminal set" is defined as a set of states such that once the network is in any state of the set, it will never go to a state which is not one of the states of the set. For example, in Table I(A) states 3 and 4 form a terminal set.

If any output Z_k is equal to 1 for all states of a terminal set, it is possible to generate Z_k by means of an OR-gate network. This is done by setting $z_k=1$ for all combinations of inputs and input-states which have as output-states, states which are members of the terminal set. For input states which are members of the terminal set, z_k can be either zero or one (z_k is a "don't care") since the required output has already been generated by a previous cell. Table I(B) shows the modification of Table I(A) which arises because $Z_k=1$ for the terminal set consisting of states 3 and 4. The circuit output Z_2 is derived from an OR-gate whose inputs are the z_2 outputs of the individual cells.

If an output Z_r is equal to 0 for all states of a terminal set, an AND-gate realization is possible with z_r equal to 1 for all combinations of inputs and input-states except those which have a state of the terminal set as their output-state. Again, z_r is a "don't care" when the input-state is one of the states of the terminal set.

In Table I(B), states 1 and 3 are equivalent if z_2 is set equal to 0 for $s_j=3$, and states 2 and 4 are equivalent if z_2 is set equal to 0 for $s_j=4$, $x_i=0$ and z_2 is set equal to 1 for $s_j=4$, $x_i=1$. Table I(B) can therefore be simplified to Table I(C).

In the example just discussed, no states were eliminated in forming the table for the OR-gate network. In general it is only possible to eliminate a terminal set for which all outputs have the same values for all members of the set (such a terminal set can contain only one state and is therefore a terminal state). An output-state entry becomes a "don't care" only when the corresponding input-state is eliminated from the table. This is illustrated by state 5 in Table II(A).

Table II. State Tables for Network for Which $Z_1=1$ If 4 or More Inputs Are Equal to 1, And $Z_2=1$ If 2 or Fewer Inputs Are Equal to 1

$s_j \backslash x_i$	0	1	Z_1	Z_2
1	1	2	0	1
2	2	3	0	1
3	3	4	0	1
4	4	5	0	0
5	5	5	1	0

S_j

(A)

$s_j \backslash x_i$	0	1
1	1,01	2,01
2	2,01	3,01
3	3,01	4,00
4	4,0d	—,1d

$S_j, z_1 z_2$

$Z_1 = z_{11}+z_{12}+\dots+z_{1n}$

$Z_2 = z_{21} z_{22} \dots z_{2n}$

(B)

It is possible to have a network in which one of the outputs is generated by an AND-gate and another output is generated by an OR-gate. For example, in Table II(B), Z_1 must be generated by an OR-gate and Z_2 must be generated by an AND-gate.

The preceding discussion of circuit-output networks has been concerned solely with electronic networks. For relay networks, the remarks concerning OR-gate and AND-gate networks are not applicable since there is not a corresponding delay associated with transmitting a signal through many cells. The relay circuit corresponding to an OR-gate network would be a network with an output bus wire which can have a signal placed on it from any cell. There is no relay circuit corresponding to the AND-gate network. Whether a relay network will have an output bus wire will be determined directly by the coding which is assigned to the states.

It has been shown that single-output circuit-output networks can be designed as OR-gate networks if there is a terminal state for which the output is to be one and as AND-gate networks if there is a terminal state for which the output is to be zero. For multiple-output networks the situation is the same except that terminal sets of states rather than terminal states must be considered. Examples have been given for the procedure to be followed in converting to OR-gate or AND-gate networks.

References

1. On the Inversion Complexity of a System of Functions, A. A. Markov. *Journal*, Association for Computing Machinery, Baltimore, Md., vol. 5, 1958, pp. 331–34.

2. Minimization of Boolean Functions, E. J. McCluskey, Jr. *Bell System Technical Journal*, New York, N. Y., vol. 35, 1956, pp. 1417–44.

3. A Topological Method for the Determination of the Minimal Forms of a Boolean Function, R. K. Mueller, R. H. Urbano. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., vol. EC-5, 1956, pp. 126–32.

4. Complete Decoding Nets: General Theory and Minimality, A. W. Burks, R. McNaughton, C. H. Pollmar, D. W. Warren, J. B. Wright. *Journal*, Society for Industrial and Applied Mathematics, Philadelpnia, Pa., vol. 2, 1954, pp. 201–43.

5. Iterative Combinational Switching Networks—General Design Considerations, E. J. McCluskey, Jr. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, vol. EC-7, 1958, pp. 285–91.

6. The Design of Switching Circuits, W. Keister, A. E. Ritchie, S. H. Washburn. D. Van Nostrand Company, Inc., Princeton, N. J., 1951.

7. Switching Circuits and Logical Design, S. H. Caldwell. John Wiley & Sons, Inc., New York, N. Y., 1958.

8. The Synthesis of Iterative Switching Circuits, D. A. Huffman. *Quarterly Progress Report*, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Mass., Jan. 1956, pp. 63–67.

9. The Synthesis of Sequential Switching Circuits, D. A. Huffman. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 257, 1954, pp. 161–90.

10. Representation of Events in Nerve Nets and Finite Automata, S. C. Kleene. "Automata Studies," C. E. Shannon, J. McCarthy, editors, Princeton University Press, Princeton, N. J., 1956, pp. 3–41.

11. Realization of Events by Logical Nets, I. M. Copi, C. C. Elgot, J. B. Wright. *Journal*, Association for Computing Machinery, vol. 5, 1958, pp. 181–96.

12. Sequential Circuit Theory, W. J. Cadden. *Ph.D. Thesis*, Princeton University, May, 1956.

13. See reference 7, chap. 15.

The K24A Syncroplex Telephone Carrier System

B. G. COETSEE
NONMEMBER AIEE

G. L. CURTIS
MEMBER AIEE

J. W. HALINA
NONMEMBER AIEE

Part I: Historical Background and Description

A MOST significant turning point in the history of telephone transmission occurred shortly before 1920 with the introduction of carrier equipment as a means of multiplying the capacity of open-wire line facilities. The 20-year period between 1918 and 1938 witnessed the firm establishment of single-side-band techniques in telephone transmission through the *C* carrier and the development of the 12-channel *J* and *K* systems for open wire and cable, respectively. This was a period of rapid expansion of carrier applications to long-haul toll circuits and, by and large, a consolidation of carrier design for this realm of service. If the *L* system is included, which is essentially an expansion of the *K*, the equipment which supports the majority of long-haul toll transmissions is still of the *J*, *K*, or *L* type.

A new departure in carrier equipment development occurred around 1950 with the introduction of carrier to short-haul toll circuits, as well as to tributary, trunk, and long-exchange lines. New economies in design and manufacture produced the *N*, *O*, and *ON* equipments, a class of systems which made the application of carrier to circuits between 20 and 200 miles long economically feasible. With the advent of these carrier systems, the number of carrier channel miles in the nation advanced to a new order of magnitude.

As this application of carrier began to gain momentum, the attention of telephone communication engineers turned to an area of the telephone transmission plant which had heretofore defied the economic application of carrier techniques on any significant scale, viz., the trunk and exchange circuit where more than 50% of the circuits fall in the 5- to 20-mile range. If a carrier equipment could be produced to serve this class of circuit, still another order of magnitude of carrier application to telephone transmission could be achieved, with its attendant gain in service to the subscriber.

After a study of design possibilities and available techniques, a decision was made at the ITT (International Telephone and Telegraph Corporation) Laboratories in mid-1957 to proceed with the development of an exchange carrier system. This culminated in the field trials of the *K24* Syncroplex System towards the end of 1958. Because of the tremendous pressure of demand for circuits, an objective was set up to compress the usual 4- to 5-year cycle of development and testing of a basically new carrier system into $1^{1}/_{2}$ to 2 years.

Characteristics of Trunk and Exchange Transmission Facility

One of the most important aspects of the design and development problem in the telephone transmission field is that of a large existing facility which it is desirable to improve and which it is virtually impossible to redesign in any radical way. This, then, means the development of equipment to work on, and with, existing facilities which were frequently designed and installed at great cost, for a different type of use than that now intended. If one were starting over again, more suitable facilities could be designed, but this is small comfort.

In the case of the exchange and trunk facility, there is an essentially low-frequency medium. For voice transmission, these lines are equalized for response in the 0- to 4-kc frequency range by the addition of loading coils. Repeaters of the hybrid or negative-impedance type are employed to provide gain where necessary. In addition to voice, the physical line is, of course, capable of transmitting d-c signaling pulses and all possible states of the loop battery; i.e., battery on, off, normal, or reverse. If a carrier system were economically to replace this service, it would have to be capable of using the facility to an order of 1/2 mc (megacycle), accommodate a relatively complicated signaling transmission requirement, provide a very high degree of quality and reliability, be simple to install and maintain, and cost one half to one quarter the cost of previously applied systems, a truly imposing challenge.

The electrical characteristics of typical exchange lines are depicted in Fig. 1. The attenuation characteristics are shown

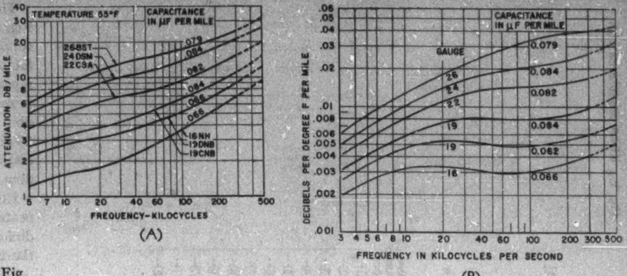
Vol.
Pt. 1
1959

Paper 59-1104, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted June 8, 1959; made available for printing August 20, 1959.

B. G. Coetsee, G. L. Curtis, and J. W. Halina are with International Telephone and Telegraph Corporation, Palo Alto, Calif.

Fig. 1. Typical exchange lines

A—Attenuation characteristics
B—Thermal characteristics



(A)

(B)

in Fig. 1(A) and the thermal in Fig. 1(B).* Three other characteristics of great importance are those of near-end crosstalk, far-end crosstalk, and impulse noise power. These are parameters which, due to the nature of the medium, can only be described statistically. On the basis of very limited tests, the range in which these parameters fall is shown in Table I.

From an assessment of this data, it follows that an exchange carrier system must have the following principal characteristics:

1. Economical application on line lengths between 5 and 20 miles, or attenuations in the order of 200 db (decibels) at 500 kc.
2. Repeaters providing for equalization in the low-frequency to 500-kc range of from a few db to a maximum of approximately 16 db.
3. Adequate system regulation for 15- to 18-db variation in 175- to 200-db lines.
4. Transmission quality of a maximum of 20 dba (db adjusted) for all types of noise.

The fourth requirement implies that for a 15-mile system the minimum signal levels would have to stay above −35 dbm (db referred to 1 mw[milliwatt]) (lower, to the extent that the modulation system chosen would have some improvement factor). In addition, the system would have to provide for 15 to 25 db of improvement on far-end crosstalk for about 25% of the pairs and 10 to 20 db of improvement on near-end crosstalk, again for about 25% of the pairs, depending on level differentials between the maximum transmitting power and the minimum receiving power.

In consideration of power and space economies, as well as design obsolescence, it was a foregone conclusion that design would be based on transistors rather than on vacuum tubes.

Choice of Modulation System

One of the obvious answers to many of the transmission problems would have been the choice of a modulation system capable of trading noise for bandwidth, such as pulse code. However, adequate data were not readily available with regard to the peak impulse interference characteristics of exchange cables, nor could such information be gathered fast enough to assure the development of a system in the time prescribed. On the other hand, consideration of the near-end crosstalk and available power-handling capacities of high-frequency transistors indicated that a modulation scheme which was reasonably conservative of spectrum consumption and reasonably efficient with regard to multichannel loading would be desirable. A modulation system which did not require a wide band per channel would permit an interleaved modulation spectrum for near-end crosstalk immunity without going beyond 500 kc of spectrum for the 24 channels. If, in addition, it did not emit a redundant and power-consuming carrier, it would be possible to make efficient use of readily available transistors which offered a peak power-handling capacity of the order of +13 dbm.

At about this time, a development at the ITT Laboratories produced an economical means of detecting double-sideband suppressed-carrier amplitude modulation (a-m). Although it was recognized for many years that double-sideband suppressed-carrier a-m could offer a close approach to the performance characteristics of single-sideband suppressed-carrier a-m in the matters of band conservation and load factors while avoiding the high channel filter costs of single sideband, its use had been hampered by the lack of a means of detection sufficiently economical to warrant its choice over single sideband in consequence of its use of twice the bandwidth and its fourfold greater loading factor. In view of the laboratory development of a detector, it appeared that a modulation system was available to satisfy all but one of the combined requirements of an exchange system, viz., far-end crosstalk on about 40 or 50% of arbitrarily selected pairs. It was decided that the system would be so designed that a compandor or some other type of noise-reducing circuitry could be introduced by the simple substitution of a plug-in panel. In this manner, a system was available which, for 50 to 60% of the arbitrarily chosen pairs, would not require the expense of noise-suppression equipment, while, at the same time, the infrequent incidence of pairs with higher crosstalk would permit simple treatment by the substitution of the vf (voice frequency) panel by an alternate one containing circuitry furnishing the necessary noise suppression. A Sonad* (Speech-Operated Noise-Attenuation Device), providing an additional 14–18-db noise suppression during idle-circuit periods, was developed at the ITT Laboratories and is undergoing extensive tests to determine its application possibilities. For larger improvements, 20–26 db, a syllabic compandor is available.

Table I. Cable Crosstalk Characteristics

	Per Cent of Pairs Whose Characteristics Are Worse Than Value Indicated			
Parameter, 500 Kc	100%	50%	25%	10%
Near-end crosstalk coupling loss, db	95	82	74	68
Far-end crosstalk coupling loss, db/mile	85	63	56	52
Impulse noise, dbm, rms	−100			

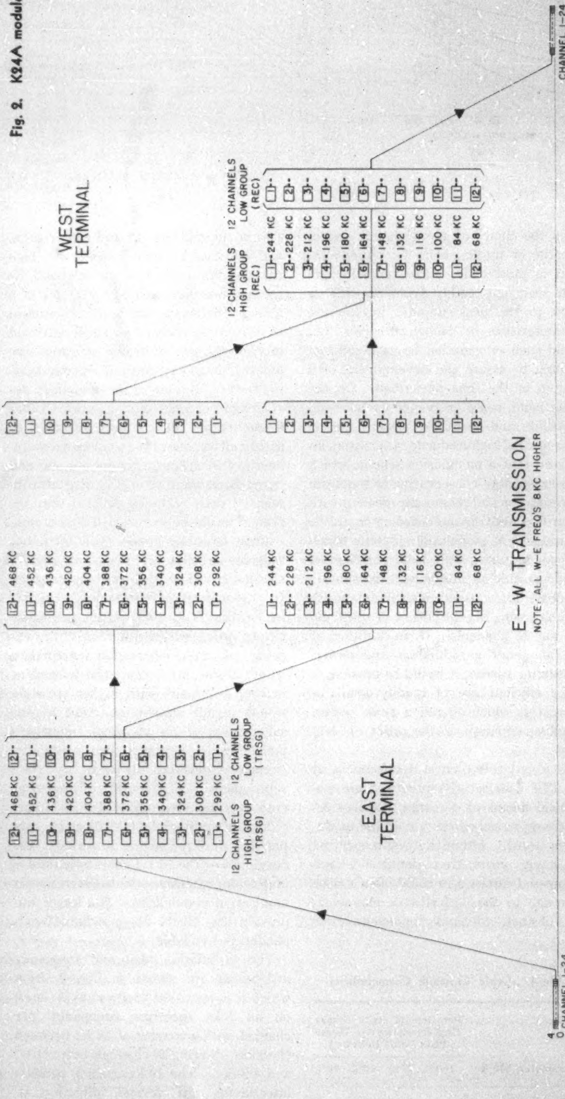
The modulation plan and frequency assignment are shown in Fig. 2, from which it is seen that the system is based on an 8-kc spectrum occupancy per channel, with a spacing of 16 kc between channels, allowing 24 channels between 60 and 480 kc. The 16-kc spacing permits interleaving the reverse direction of transmission and thus provides for immunity to near-end crosstalk.

---

* Figs. 1(A) and (B) were reproduced from Figs. 1 and 6 of reference 1 by permission. Extrapolations for frequencies beyond 250 kc are the responsibility of the authors.

* Patent applied for.

Fig. 2. K24A modulation plan


WEST TERMINAL
EAST TERMINAL
12 CHANNELS HIGH GROUP (REC)
12 CHANNELS LOW GROUP (REC)
12 CHANNELS HIGH GROUP (TRSG)
12 CHANNELS LOW GROUP (TRSG)
468 KC
452 KC
436 KC
420 KC
404 KC
388 KC
372 KC
356 KC
340 KC
324 KC
308 KC
292 KC
244 KC
228 KC
212 KC
196 KC
180 KC
164 KC
148 KC
132 KC
116 KC
100 KC
84 KC
68 KC
E – W TRANSMISSION
NOTE: ALL W–E FREQ'S 8KC HIGHER
CHANNEL 1–24
4
0

The loading characteristic is shown in Fig. 3 in comparison with single-sideband suppressed carrier. With a 20-db load factor for 24 channels and a transistor-repeater overload capacity of +13 dbm, it was possible to transmit at a level of −8 dbm per channel and handle a better than 25-db span without dropping below the established minimum signal level of −35 dbm.

Regulation was incorporated in the repeaters to provide for a stability of ±1 db for an 8-db line-level variation; and in the channel terminals, for a stability of ±0.5 db for an 8-db variation, resulting in a level stability of ±0.5 db for a 175-db system. Regulation is not affected by the addition of SONAD circuitry. In the case of compandored channels, the stability is 1.0 to 1.5 db under the same line conditions.

Equipment linearity was maintained to hold all types of distortion down to 20 dba for an 8-repeater system.

The system block diagram is shown in Fig. 4.

Signaling

An exchange carrier system must provide a choice of signaling facilities not previously provided as an integral part of carrier systems. Two types of signaling are of principal interest. In their simplest terms these two types require the transmission of the following information:

1. *Loop signaling.* This signaling option requires the transmission of closed-loop open-loop conditions from an originating to a terminating office at a rate of 0- to −14 alternations per second. In the reverse direction, the transmission is that of the condition of the battery, normal or reverse, at a somewhat slower rate.

2. *Revertive signaling.* In this signaling option, the originating-to-terminating office transmission consists of loop-open loop-closed conditions at up to an equivalent of 36 alternations per second and, in the reverse direction, all states of the loop battery; viz., on, off, reverse, or normal.

The familiar E and M (electric and magnetic) signaling facilities normally offered by carrier systems must be augmented by a 4-wire to 2-wire, or a 4-wire to 3-wire converter, to provide loop signaling. The third wire in the 3-wire option furnishes supervisory information normally provided by the outgoing trunk circuit and, when used, eliminates the need for this appliqué panel. In the case of the revertive system, a facility capable of transmitting one-half bit more information from the terminating office to the originating office is necessitated, a requirement of one-half bit more than

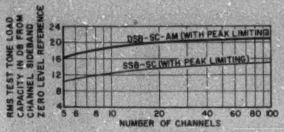


Fig. 3. Comparison of load factors, DSB-SC versus SSB-SC

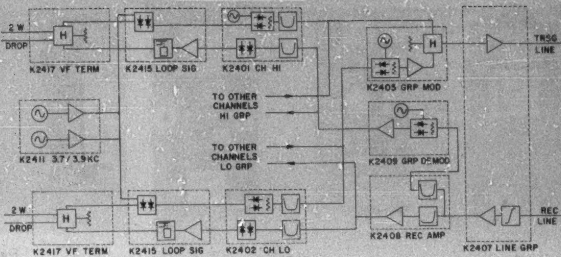


Fig. 4. K24A terminal

E and M or loop-signaling systems are capable of handling. Despite the fact, therefore, that a double-sideband suppressed-carrier a-m system offers an economical one-bit transmission system in the availability of the carrier for control-information transmission purposes, the need for providing for revertive signaling led to the choice of frequency-shift signaling, which is capable of two bits of information transmission by offering the alternatives of sending either one of the two frequencies, both, or none at all. Moreover, it yields a vehicle for channel regulation.

The resulting signaling facility is thus, in fact, an excellent telemeter channel, capable of handling up to 30 bits per second with pulse distortion less than 5%.

Mechanical Aspects of K24 System

In a search for the simplest and most flexible mechanical formula, a computer-type system was evolved, as shown in Fig. 5. The system features printed-circuit wiring, "banjo"-type rack wiring and plug-in, "bookshelf" construction. As indicated earlier, a channel is broken up to permit the easy assembly of a system for any application; signaling cards can be readily substituted to provide E and M, loop, or revertive signaling without disrupting or realigning the remainder of the system. Similarly, the vf term panel can be readily substituted to provide 2-wire, 4-wire, compandor, or SONAD facilities, etc.

SUMMARY OF SYSTEM PARAMETERS

Following is a summary of the *K24* system parameters:

Number of channels, 24
Frequency spectrum, 60 to 480 kc
Transmit levels per channel sideband, −8 dbm
Receive levels per channel sideband, −33 dbm
24-channel load factor, 20 db
Repeater maximum power capacity, +13 dbm

Span: attenuation, 25 db maximum
slope, 18 db maximum
System regulation, ±0.5 db
Channel equivalent, 0 or −3 db
Terminal noise, 20 dba at the zero level (maximum all types)
System noise, 25 dba maximum for 175 db
Signaling, E and M, loop, or revertive
Compandors, optional
SONAD, optional
Size, 31.5 inches of rack space per 24-channel terminal
Power consumption, approximately 100 watts at 48 volts per 24-channel terminal
Repeaters, power consumption 50 mils at 30 volts; optional, regulated or non-regulated
Alarms, power failure, common equipment failure, and system failure

Part II: Equipment Considerations

The *K24A* Syncroplex Telephone Carrier System is a 24-channel multiplex equipment based on double-sideband suppressed-carrier modulation, in which a base group of 12-channel modulators (transmitters) operates at carrier frequencies between 292 and 476 kc and a corresponding base group of 12-channel demodulators (receivers) operates between 68 and 252 kc. Each channel occupies 8 kc of spectrum; thus adjacent channel spacing can be as close as 8 kc. In the case of the *K24A* system, which is arranged for exchange cable applications, the spacing is 16 kc to permit interleaving of the two directions of transmission and to prevent crosstalk as explained in Part I.

To close the transmission path between the 12-channel modulator group and the 12-channel demodulator group, it is thus necessary to translate (group-modulate, or demodulate) once.

For an appreciation of the equipment design, the *K24A* system can be divided into four parts; namely:

1. *Channel terminal equipment.* This group, shown in Fig. 6, consists of the modem unit necessary to derive the channel signal from the multichannel spectrum, the signaling unit necessary to accommodate the transmission of various systems of switching information, and the terminating unit necessary to provide 4-wire to 2-wire conversion for the signaling circuit, as well as other functions such as vf companding, signal limiting, etc.

A great variety of arrangements of chan-

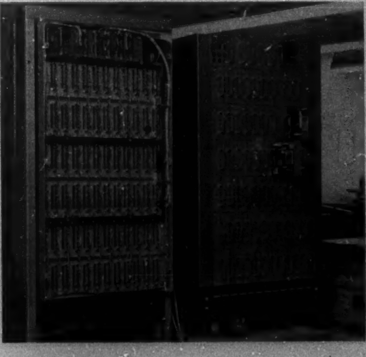


Fig. 5. Prototype of K24A terminal equipment

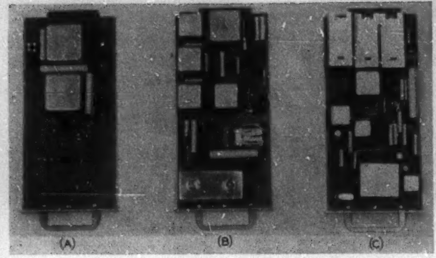


Fig. 6. K24A channelizing equipment

A—Vf terminal
B—Loop signal, originating office
C—Channel modem

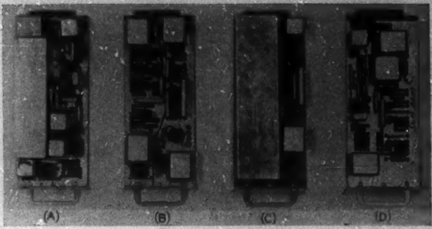


Fig. 7. K24A group equipment

A—Group modulator
B—Line amplifier
C—Receiving amplifier
D—Group demodulator

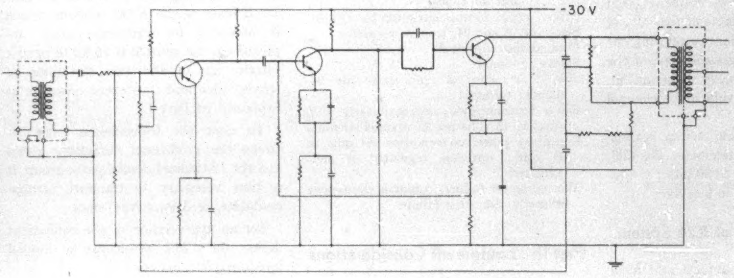
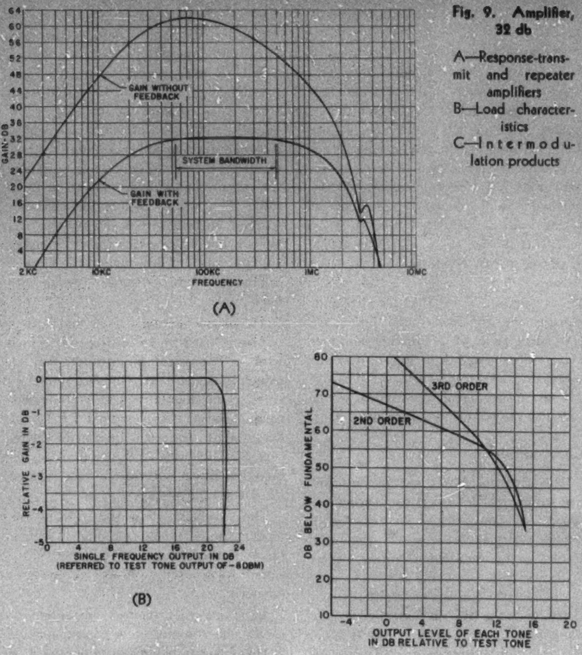


Fig. 8. Simplified schematic of 32-db amplifier

Table II. Functional Description of K24 Units

Group	Unit	Option	Function	Type No
I. Channel terminal equipment	Channel modem	Master	Channel transmitter and receiver including crystal controlled carrier supply for two transmitters	K2401
		Slave	Like K2401, except without carrier oscillator	K2402
	Signaling unit	E and M	Frequency-shift transmitter and receiver discriminator, including signaling tone pick-off filter, agc rectifier, and relay	K2419
		Loop, 2-wire	Like K2419 except arranged to operate from and into a-b leads of vf term set. Provided in originating and term office versions	K2415–16
		Loop, 3-wire	Like K2415–16 except control lead added at originating office to eliminate need for outgoing pulse repeaters in step-by-step offices	K2425–26
		Revertive	Like K2415–16 except equipped to provide channel for revertive pulse transmission	K2403–04
	Vf term unit	2- to 4-wire term set	2- to 4-wire vf hybrid, a-b lead connections, vf limiter	K2417
		2- to 4-wire term set and Sonad	Like K2417 except Sonad added	K2418
		2- to 4-wire term set and compandor	Like K2417 except compandor added	K2428
II. Terminal common equipment	Group modulator		Translates 12 high group transmit channels to low group spectrum, combines both groups, generates pilot tone	K2405
	Group receiving amplifier		Separates high group from low group, amplifies low group for channel detection	K2408
	Group demodulator		Translates 12 high group received channels to low group spectrum and amplifies the group for channel detection	K2409
	Line amplifier		Amplifies channel to line level for transmission, provides span adjustment, equalization and amplification for received signals. Includes facilities for connecting repeater power	K2407
III. Terminal auxiliary equipment	Signal supply		Low-impedance power source of 3.7-kc and 3.9-kc tones for 24 signaling units	K2411
	Fuse and alarm panel		Contains fuses, power-supply failure alarm, pilot failure alarm, common equipment and channel failure alarm	K2406
	Power supply regulator		Converts −48 volt close-open supply to −30-volt regulated supply	K2412
IV. Repeater equipment	Repeater	Regulator repeater	25-db span amplifier including span pads and equalizers	K2414
		Nonregulator repeater	Like K2414 but without regulating circuit	K2413
V. Terminal shelf			Accommodates the 79 panels composing a 24-channel terminal, terminal strip, and repeater supply current control	K2410
VI. Test, line-up and maintenance equipment	Alignment set		Facilities for testing repeaters and line sections for span length adjustment, equalization, and trouble shooting	K2480
	Unit-extension test sets		Facilities for extending units out of shelf and making test measurements	K2481–3



(A)

(B)

(C)

Fig. 9. Amplifier, 32 db

A—Response-transmit and repeater amplifiers

B—Load characteristics

C—Intermodulation products

nel terminal is possible through the appropriate choice of signaling and vf terminating units, thus yielding excellent flexibility for many system applications.

2. *Group equipment.* The set of units shown in Fig. 7 is comprised of the 12-channel group-frequency translating equipment and a 24-channel line amplifier.

3. *Auxiliary equipment.* This group includes the terminal-voltage regulator, the fuse-and-alarm unit, and the 24-channel signaling-tone supply.

4. *Line equipment.* For operation over exchange cables, there are provided repeaters of both regulating and nonregulating types, equipped with span pads and equalizers.

For ease of reference, the entire complement of units and their functional descriptions are summarized in Table II.

Circuits

The *K24A* system uses transistors throughout the equipment. This represents one of the first large-scale applications of transistors to carrier telephony up to 1/2 mc of carrier spectrum. Many stringent performance requirements, which are probably more peculiar to multichannel telephony than to other areas of communication, have to be met. Accordingly, several basic circuits were developed and were used in a variety of functional units. The principle of these units will be described now and later referred to in the description of units in which they occur.

AMPLIFIER, 32 DB[2–4]

This is a 3-stage 60–480-kc wide-band amplifier with an average of 25 db of feedback in the specified range of application. An interesting aspect of this unit is that the transistors used in it exhibit a mean beta cutoff of 250 kc, whereas the amount of feedback necessary to achieve linearity required a controlled feedback characteristic to beyond 4 mc.

A simplified schematic of this amplifier is shown in Fig. 8. It consists of three common emitter stages in cascade. The design was accomplished by successive *R-C* compensation of each of the three stages in such a way as to maintain a 30-degree phase margin to beyond the point of unit gain in the feedback loop. The output impedance is bridge stabilized, and the input impedance is stabilized by means of passive resistance. The result is an over-all impedance stability of better than 5% across its operating band.

The feedback and unfeedback gain characteristics of the amplifier are shown in Fig. 9(A), its load characteristic in Fig. 9(B), and its linearity characteristic in Fig. 9(C).

The amplifier is the basic circuit in the regulating and nonregulating repeaters, the transmit section of the line amplifier, and the group demodulator.

AMPLIFIER, 15-DB

Except for gain, the characteristics of this amplifier, insofar as linearity, feedback, and load-handling capacity are concerned, are the same as those of the 32-db amplifier. The lower gain required permits the use of two stages, a common base and a common emitter, in cascade. The amplifier is impedance stabilized with bridge feedback at both the output and the input.

CRYSTAL-CONTROLLED CARRIER-SUPPLY OSCILLATOR

The requirement for stable medium-power constant-voltage source of carrier supply exists in many units of the system, e.g., channel modulators, group modulators and demodulators, pilot generators, etc.

The circuit chosen was that of a modified Colpitts[5] (commonly referred to as the Clapp or Gouriet) oscillator, d-c coupled to a boot-strap amplifier. A simplified circuit is shown in Fig. 10.

The use of a boot-strap buffer provides a device with a very high impedance load for the crystal oscillator, a very low output impedance which permits driving multiple loads with extremely low coupling, and good power gain and overload characteristics. The resultant carrier supply is capable of 20–24 mw of output power from a source impedance of a few ohms.

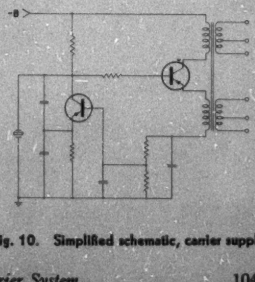


Fig. 10. Simplified schematic, carrier supply

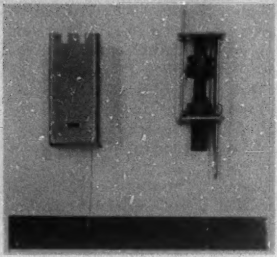


Fig. 11. K24A channel transmit filter

L-C Oscillator

Due to the cost or size of low-frequency crystals, or the inflexibility of crystal oscillators, a stable *L-C* circuit is desired in units such as the signaling-tone supply and the local oscillator source in the channel modem. For these applications, a tuned collector-emitter feedback circuit was designed. The oscillator operates at frequencies of 3.7 to 252 kc and, without the aid of a buffer, provides over 25 mw of power with a frequency stability of better than 0.1% over a temperature range of 0–60 C (degrees centigrade).

Filters

Considerations of economy in filtering are all-important in carrier equipment design, particularly at the channel-modem level where they are numerous. The structure used in the *K24A* system is that of an intermediate-frequency-type filter in which universal coil construction and adjustable leakage coupling are used to achieve the elements of a mesh. Fig. 11 shows the transmit filter, and Figs. 12(A) and (B) are typical insertion-loss curves of the transmit and receiver filters.

The simplicity of the receive filters results in low selectivity, but this can be tolerated because both the adjacent channels are demodulated out-of-band with synchronous detection; in single sideband, with normal channel spacing, one of the adjacent channels will be demodulated in-band. With emitted-carrier double-sideband and envelope detection, all channels are demodulated in-band.

Units of K24A System

The functions of the various units constituting a *K24* terminal, briefly described in Table II, are further described in the block diagram of Figs. 13(A) and (B).

Units of particular interest will be detailed in the following paragraphs.

Channel Modem

The carrier oscillator used in the transmit section was described in an earlier section. The remainder of the circuit consists of two passive elements, a balanced lattice modulator, and a transmit filter of the type shown in Fig. 11.

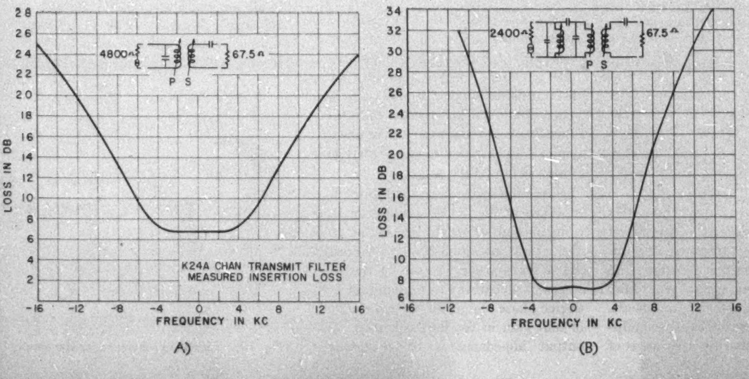
The receiver consists of essentially four elements: a variolosser controlled by the level of the received frequency-shift signal in the signaling unit, a carrier amplifier, a local oscillator, and a ring demodulator.

The ability of the receiver to detect double-sideband suppressed-carrier a-m signals is due to a synchronizing circuit which controls the local oscillator. With the synchronizing circuit disabled the unit resembles a conventional nonsynchronous single-sideband demodulator. In the case of the reception of double-sideband suppressed-carrier a-m signals, the synchronizing circuit detects the zero crossings of the double-sideband channel signal. These occur at twice the suppressed-carrier rate, with perturbation at a rate equal to twice the modulating frequency. The local oscillator is designed to synchronize at alternate impulses provided by the synchronizer and to ignore the perturbing impulses due to modulation. The result is a local oscillator source, synchronized rigidly in both frequency and phase, which therefore makes possible coherent detection of the double-sideband suppressed-carrier signal with fidelity.

The receive section of a channel modem is thus a fixed-frequency receiver which with an appropriate transmitter can, in principle, detect double-sideband emitted-carrier a-m, double-sideband suppressed-carrier a-m, or single sideband. To the extent that a carrier is present in the received double-sideband signal, it is detected as direct current and can be used for control purposes.

The excellent phase linearity of the modem can be judged from Fig. 14, which displays the response of the modem to square-wave excitation at various frequencies.

As noted earlier, the transmitter and receiver, on a common modem card, operate in two separate bands of frequencies. A number of considerations led to this arrangement. High-frequency operation of transmitters permits an economical crystal oscillator design. On the other hand, the receiver is better operated at low frequencies due to improved performance of transistors. In addition, of course, the wide separation of frequencies obviates any coupling problems between transmitter and receiver on a single card.



A)

(B)

Fig. 12. K24A channel filter insertion-loss curves

A—Transmit filter
B—Receive filter

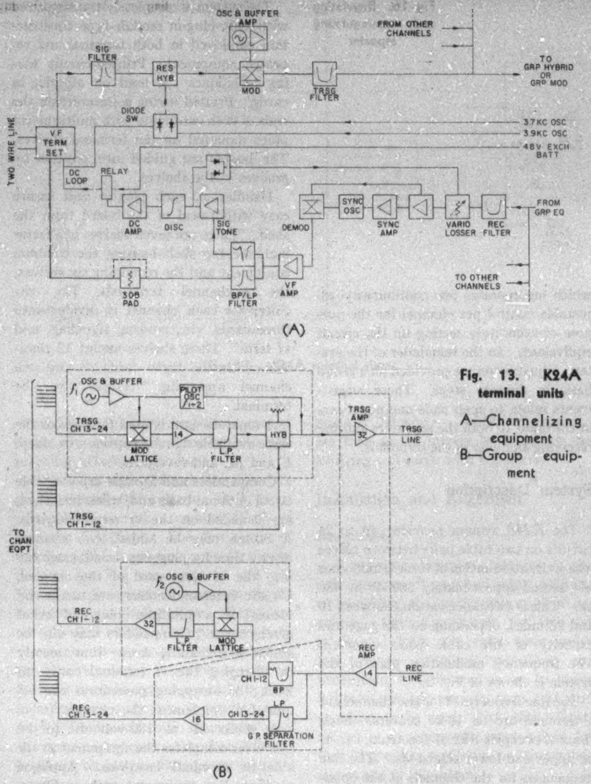

OSC & BUFFER AMP
SIG FILTER
RES HYB
MOD
TRSG FILTER
FROM OTHER CHANNELS
TO GRP HYBRID OR GRP MOD
DIODE SW
3.7KC OSC
3.9KC OSC
48V EXCH BATT
TWO WIRE LINE
VF TERM SET
DC LOOP
RELAY
DC AMP
DISC
SIG TONE
DEMOD
SYNC OSC
SYNC AMP
VARIO LOSSER
REC FILTER
FROM GRP EQ
30B PAD
BP/LP FILTER
VF AMP
TO OTHER CHANNELS
(A)
OSC & BUFFER
f_1
PILOT OSC $f_1 \div 2$
TRSG CH 13-24
MOD LATTICE
14
L.P. FILTER
HYB
TRSG AMP
32
TRSG LINE
TRSG CH 1-12
TO CHAN EQPT
f_2
OSC & BUFFER
REC CH 1-12
32
L.P. FILTER
MOD LATTICE
REC AMP
14
REC LINE
CH 1-12
BP
LP
REC CH 13-24
16
CH 13-24
G.P. SEPARATION FILTER
(B)

Fig. 13. K24A terminal units

A—Channelizing equipment

B—Group equipment

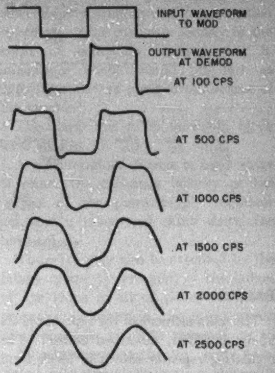


Fig. 14. Channel response to square-wave input

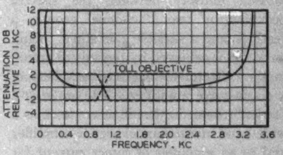


Fig. 15. Channel frequency-attenuation characteristics

SIGNALING UNIT

There are a number of systems of transmission of switching information between exchanges. The simplest, and perhaps most common, is E and M, which is essentially the provision of a 4-wire low-speed data-transmission channel in conjunction with the voice circuit. Two other systems are usually referred to as "loop" and "revertive." In loop signaling, the the 4-wire E-and-M data circuit is converted to 2-wire operation over the loop used for voice transmission. An optional version provides supervisory information on a third control wire, eliminating the trunk circuit normally required. In revertive signaling applications, the functions of loop signaling are augmented by a third supervisory function. Space does not permit a detailed description of these systems.

In order to satisfy the functional requirements of these three systems of signaling transmission, three options of signaling unit are available as shown in Table II.

The need for transmission of more than one bit of information in the case of revertive signaling influenced the choice of frequency-shift modulation for the signaling channel, in that three states are possible: one or the other of the two frequencies, or, for short duration, both.

Transmission of a frequency-shift signal is accomplished by the provision of two tones to all 24 channels from a common supply. The tones are at 3.7 and 3.9 kc, just above the voice band. The "shift" is achieved in the signaling transmitter by means of a diode switch designed to fade-in and fade-out the alternate frequencies and thus hold fortuitous switching distortion to acceptable limits. The signaling receiver is a conventional Round-Travis discriminator which serves the dual purpose of driving a signaling relay and an agc (automatic gain control) rectifier.

Three steps of signaling distortion adjustment are provided to compensate for varying external plant conditions.

VF TERMINATING UNIT

This unit contains as its main element a transformer bridge, or hybrid, which serves to convert the 4-wire voice circuit provided by the modem to a 2-wire circuit capable of connecting with the exchange loop. It similarly permits the conversion of the 4-wire facility of the signaling unit into loop signaling.

The SONAD has been designed so that it may be added to the basic vf terminal card at any time. An optional version of the vf terminating unit contains a compandor, thus permitting the application of the system to exceptionally noisy transmission media.

Fig. 15 indicates the channel-response characteristics. It is largely influenced at the lower end of the vf spectrum by the characteristics of the terminating unit and, at the upper end, by the signaling-tone reject filters in the signaling unit.

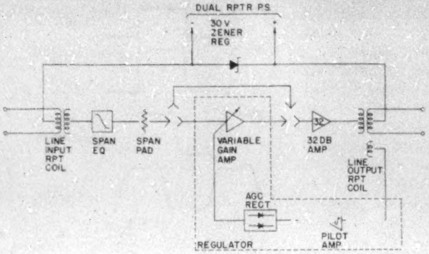


Fig. 16. Regulating and nonregulating repeater

REPEATERS

The main element of the repeater is its 32-db amplifier, described earlier. The regulating repeater shown in block form in Fig. 16 contains an additional circuit. This circuit consists of a frequency-selective amplifier which picks off a pilot located midway between the spectra occupied by two 12-channel groups, a rectifier, and a variable gain stage. Gain variation is achieved by thermistor action in the collector-to-base feedback loop of a single-stage amplifier. The stage is a "reflex" circuit in that it serves both as a wide-band multichannel signal amplifier and as an amplifier of direct current for the control of its thermistor feedback circuit.

The repeater also accommodates, as plug-in elements, a span pad and a span equalizer. Power is obtained from a constant-current 60-ma (milliampere) source which is series-fed around the phantom circuit formed by the east and west pairs of a system.

Part III: Applications and Maintenance

At the outset of the development of the *K24* system, it was determined that the success of a 5- to 15-mile short-haul carrier would be very dependent on the simplification of application, installation, lineup, and maintenance techniques to a degree not previously attained in the common carrier field. How the system meets these objectives is discussed here, the equipment approach to the system having been outlined in Part II.

The high density of circuits used in the exchange area of telephone plant, together with the obvious impracticability of having adequate staffs of skilled personnel available for its installation, adjustment, and maintenance, dictated the necessity of eliminating measurements and calculations normally associated with a carrier installation. The end product is a system which incorporates one continuously adjustable control per channel for the purpose of accurately setting up the circuit equivalent. In the remainder of the system, adjustments are provided on a group basis in discrete steps. These adjustments relate to span pads and equalizers at the input to both the repeaters and the receiving amplifier of the terminals.

System Description

The *K24A* system provides up to 24 circuits on two cable pairs between offices where the attenuation of these trunks does not exceed approximately 200 db at 480 kc. This is a distance varying between 10 and 20 miles, depending on the gage and capacity of the cable pairs involved. The frequency modulation plan of the system is shown in Fig. 2.

The line frequencies for the channels of a terminal are on 16-kc centers. Each channel occupies 8 kc of spectrum, i.e., 4-kc upper and lower sidebands. The line frequencies for the channels at the opposite terminal are also on 16-kc centers, but all frequencies are shifted 8 kc above those for the first terminal. Since each channel occupies 8 kc of spectrum, there is no overlapping of frequencies when considering the two directions of transmission. This "staggering" of frequencies for opposite directions of transmission eliminates near-end crosstalk problems and simplifies filter requirements.

The lowest channel-carrier frequency in the top 12-channel group is 292 kc, and the highest channel-carrier frequency is 476 kc.

The line frequencies of the high group were chosen to permit operation of 12-channel systems in the same cable sheath with Western Electric Company *N-1* carrier systems.

The system terminal is designed to mount in $31^1/_2$ inches of vertical space on 19-inch relay racks. Four system terminals will mount in one 11-foot, 6-inch bay.

The system is completely transistorized with unit plug-in module-type construction employed in both terminal and repeater equipment. Printed-circuit wiring techniques are used on all plug-in cards. Printed wiring connectors on the ends of each card mate with multicontact plugs mounted on the terminal shelves. The boards are guided into position by grooves in the shelves.

Handles on each plug-in unit assure easy withdrawal of the board from the shelf. There are seven shelves in a terminal, the top shelf housing the common equipment and the remaining six shelves, the 24-channel terminals. The circuitry for each channel is divided into three cards, viz., modem, signaling, and vf term. Three shelves mount 12 channels, with the three cards of any one channel appearing vertically in the terminal.

A separate card is used for each of the three available signaling options, i.e., loop, E and M, and revertive.

Compandors and SONADS are available on an optional basis and, when furnished, are included on the vf terminal cards. A SONAD may be added to a channel at any time by plugging the subassembly into the vf term card of the channel. Circuit interconnections are made by means of a multipin plug-and-socket arrangement. Compandors may also be added to a channel at any time, merely by changing the vf terminal card; no wiring or strapping procedures are required to implement the change.

A static 48- to 130-volt d-c to d-c converter completes the equipment available at terminal locations. Appliqué auxiliary line equipment such as filters, terminations, matching transformers, etc., normally associated with carrier applications, are not required when using this equipment. For instance, when the loop signaling option is used, step-by-step equipment, or the outgoing trunk circuit (depending on whether the 3-wire or the 2-wire loop signaling unit is to be used), connects directly into the drops

Table III. K24A Repeater Spacing

Cable		Repeater Spacing, Feet
Gage	Type	
16	Toll	12,000
16	NH	12,000
19	Toll	12,000
19	DNB	6,000
19	CNB	6,000
22	Toll	6,000
22	TS	6,000
22	CSA	6,000
24	NH, 0.065	6,000
24	CSH, 0.072	6,000
24	DSH, 0.084	4,500

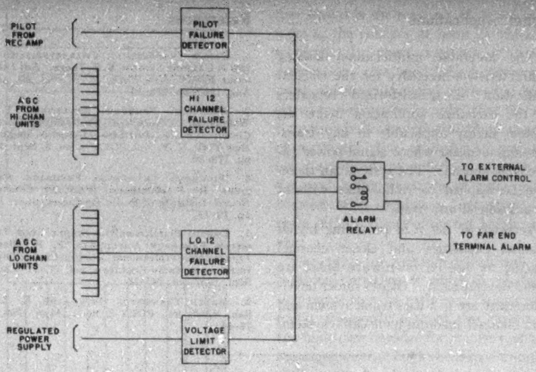


Fig. 17. K24A terminal alarm

of one terminal, and the opposite terminal drops are connected directly into the incoming selector equipment.

Regulating and nonregulating repeaters are used with this system. Both types of repeaters furnish a normal 25 db of gain at the highest line frequency of 480 kc. The regulating repeaters compensate for line variations of ±8 db. Every third repeater in a system should be of the regulating type. Repeaters are installed at the same point for both directions of transmission, regardless of the type required. Repeaters are power-fed over the same pairs on which they are installed, and up to five, in tandem, may be fed from each end. A system may use up to ten repeaters, therefore, without requiring an intermediate power source.

System Engineering

A minimum of engineering is required to install this system. At terminals, it is only necessary to specify the number of channels and the type of signaling option desired. The number and type of repeaters are determined by referring to Table III. It will be noted repeaters are normally located at 6,000-foot (*H*-loading) intervals, or multiples thereof, depending upon the gage and capacity of the cable pairs involved.

If vf-loaded pairs have to be unloaded for an initial installation of this carrier system, it is possible to effect a considerable economy in the cable-splicing operations involved. This is accomplished by unloading more than the number of pairs required for initial installation. These extra pairs are then reloaded by mounting the loading coils on plug-in cards designed to fit in the repeater housings. Additional systems are then added merely by replacing the loadging-coil cards with repeaters.

Installation and Alignment

REPEATERS

A repeater-alignment test set is the only instrument required when aligning or checking the performance of a repeater. Operating power for this set is derived from the cable pair under test. When installing repeaters, the instrument is calibrated so as to easily indicate the appropriate span pad and equalizer to be used. No other adjustments or measurement are required for installing either regulating or nonregulating types of repeaters.

TERMINALS

Alignment of the terminals consists essentially of selecting the appropriate equalizer and span pad and adjusting the circuit equivalents. The alignment test set is used at the terminal to select the pads required in a fashion similar to that used at the repeaters.

Extension units provide access to units permitting measurements on both a bridging and a terminating basis.

SIGNALING

An infrequently required, but nevertheless available, step adjustment is the one for pulse distortion. By a solder-strap change provided, bias may be varied ±10% from the nominal setting.

ALARM SYSTEM

The design of the alarm system aimed at three objectives: 1. protection of the equipment from damage due to abnormal power-supply occurrences, 2. prevention of seizure of the exchange facility in the event of equipment failure, and 3. indications which would pinpoint as unambiguously as possible the source of the trouble.

The alarm system is shown in block form in Fig. 17.

A predetermined change in the voltage of either the exchange battery or the output of the power-supply regulator actuates a relay and shuts down the terminal.

The relay may also be actuated by the failure of the system pilot or the failure of agc loops in all of either the lower group of 12 channels, the upper group of 12 channels, or all 24 channels.

Measurements at the test jacks associated with the alarm system allow the trouble shooter to isolate the failure in one of the following areas:

1. Power supply.
2. Line transmission facility, including the line amplifier in the terminal, the repeaters, or the line itself. A further degree of identification of the offending unit requires the use of the alignment test set.
3. Group modulators, demodulators, and signal supply.

Protection of the exchange from tie-up due to system failure is achieved by the inherent operating characteristics of the system. The signaling conditions corresponding to channel failure are those of either battery-normal or no-battery at the originating office and loop-open at the terminating office. This assures that in the case of failure the exchange equipment and subscribers are released.

TROUBLE SHOOTING AND MAINTENANCE

Virtually complete trouble shooting can be accomplished by means of measurements made through the unit extension panels, or by the use of the alignment test set. Test points, per se, have been eliminated from the units of this equipment in the interests of economy. However, the extension panels furnish even more complete access to each unit than that normally provided by test points.

The alignment test set described earlier is equipped, of course, to provide all of the pertinent information necessary to indicate the performance of any repeater.

POWER SUPPLY

The *K24A* system operates from 30 volts direct current. The supply is derived from the 48-volt exchange battery at the terminal locations through a power regulator. The average total consumption of a 24-channel terminal is approximately 100 watts at 48 volts.

The repeater requirement is 30 volts at 55 mils. Repeaters are supplied in series, and the individual supply is Zener regulated. Accordingly, a 5-repeater system requires a source supply of 150 volts plus the voltage drop of the cable pair. Thus, for a 1-repeater system, the 48-volt exchange battery is an adequate supply. For two repeaters or more, the 48-volt and 130-volt exchange supplies must be put in series. When the 130-volt battery is not available in an exchange, the 48- to 130-volt d-c converter may be used.

Other Applications

The favorable multichannel loading characteristics necessary for the efficient utilization of transistorized repeaters in the exchange application make the system highly applicable to any transmission medium where signal-power capacity is limited. The system can therefore be applied to radio links without adaptation of any kind.

Versions of the *K24* providing higher channel capacities and closer channel spacing for use on microwave links, are under development. Others under development are a 4-wire trunk system and a 12-channel medium-haul cable system.

References

1. Cable Transmission Characteristics for the N Carrier System, **E. F. Vaege.** *Bell Laboratories Record*, New York, N. Y., vol. 30, no. 8, Aug. 1952, pp. 333–36.
2. Wideband Feedback Amplifiers, **F. D. Waldhauer.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., vol. PGCT-4, no. 3, Sept. 1957, pp. 178–89.
3. Wideband Transistor Feedback Amplifiers, **R. P. Abraham.** *WESCON Convention Record*, Institute of Radio Engineers, pt. 2, 1957, pp. 10–19.
4. Design Principles for Single Loop Transistor Feedback Amplifiers, **F. H. Blecher.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, vol. PGCT-4, no. 3, Sept. 1957, pp. 145–56.
5. Stable Transistor Oscillator, **E. Keonjian.** *Ibid.*, vol. PGCT-3, no. 1, Mar. 1956, pp. 38–44.

◆

Discussion

Arthur F. Rose (New York, N. Y.): As the authors point out, the use of carrier for providing intercity toll circuits is an old story and over the years development engineers have had as an objective a continuous reduction in cost. This has resulted generally in cheaper and better toll service. However, up to the present time the cost reductions have not been sufficient for carrier applications in the local plant to appear economically attractive. Also there has been some reluctance to introduce new techniques to the local plant people who are relatively unskilled in the electronic arts, unless the savings appeared likely to be substantial.

The new Kellogg Switchboard and Supply Company *K24* carrier system described in this paper is so low in cost that it should extend the economic range of carrier to distances well below 10 miles. In addition, the equipment itself, the installation, and maintenance procedures, are all so simple that the plant forces should accept it as merely another tool for providing telephone circuits in the short-haul range. Several installations have been in service for many weeks and the maintenance attention required has been very small.

This new system should be of great help to the telephone companies' planning engineers who are faced with the increasing tempo of telephone usage which was not anticipated a few years back. In some cases costly conduit reinforcement can be avoided and in others the uncertainties of the future can be handled without major investment until the trends have been clarified. Changes in rate structures, involving increases in extended area service, and the consequent large increase in trunk requirements can be cared for more quickly, and frequently more cheaply, by this new type of carrier than by other forms of construction.

The telephone industry should welcome this new facility and the engineers who developed it should be congratulated on their major contribution to the communication art.

Vol.
Pt. 1
195

Asynchronous Multiplexing

J. E. TAYLOR
NONMEMBER AIEE

THIS PAPER will consider theoretical limitations and capabilities of general types of asynchronous multiplexing systems for networks of the type shown in Fig. 1. The kth station of the ith channel of the network is S_{ik}. All stations belonging to a channel are in communication with each other. The purpose of the multiplexing system is to provide this communication when stations of all channels use a common transmission medium in which signals and noise mix additively. Fig. 1 adequately describes a mobile communication system, for instance, in which the individual communication paths vary widely in direction and length.

An asynchronous multiplexing system multiplexes the individual transmissions in a way which does not require synchronization between channels. Frequency division multiplex is a well-known asynchronous system and is considered here only as a comparison standard; other systems which operate asynchronously are to be discussed.

The exclusion of frequency division coupled with requirements of no synchronization and finite time duration waveforms implies the channels will transmit nonorthogonal signals.

A pair of nonorthogonal signals are such that the energy contained in one is not completely separable from that contained in the other. This is equivalent to the mathematic statement that the inner product of the two signal functions is nonzero. If orthogonality were to be maintained for all possible time shifts between two signals, then their spectral density distributions can have no common frequency components. Similarly here, channels which are defined by the multiplexing process in a manner such that

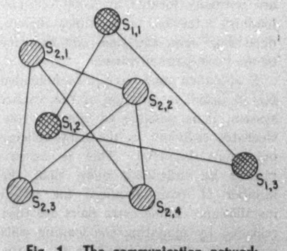


Fig. 1. The communication network

Paper 59-1146, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted January 22, 1959; made available for printing July 31, 1959.

J. E. Taylor is with the General Electric Company, Schenectady, N. Y.

The author wishes to acknowledge indebtedness to L. G. Abraham and C. A. Stutt for interesting discussions, and to R. L. Shuey, whose unfailing interest in asynchronous multiplexing stimulated most of the research reported here. The figures were prepared by J. A. Segit.

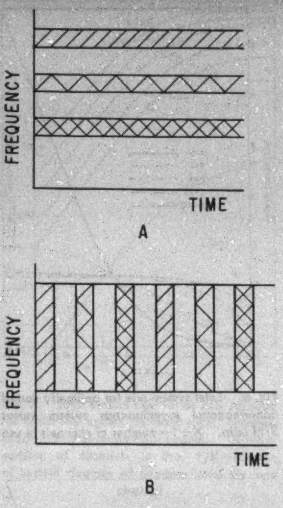


Fig. 2. Orthogonal multiplexing systems

A—Frequency division
B—Synchronous time division

signal energy in one does not get into another will also be called orthogonal. Examples of orthogonal and nonorthogonal channels are shown schematically in Figs. 2 and 3. In Fig. 2, orthogonal channels are defined by systems of frequency division and by synchronous time division, while Fig. 3 shows nonorthogonal channels which are specified by, in Fig. 3(A), a unique time period between a signal and its repetition and, in Fig. 3(B), a unique pulse repetition-frequency.

Since every reception process exhibits a threshold feature, in that a decision must be made as to what has been transmitted, it should be understood that the requirement of a communication system is to ensure almost orthogonal signals only at the place in the system where the decision must be made. Thus, in the model of a communication system shown in Fig. 4, the distinction between modulation coding and multiplex coding is made. This distinction is quite arbitrary and in some cases is difficult to make. Modulation coding determines the form in which intelligence is added to a signal, e.g., amplitude modulation. Multiplex coding is the process by which channels are separated from each other. In most communication systems the modulation coder is specified, but the input waveform is determined exclusively by the user, and it is the recovery of that waveform at the receiver at point *D* which is required. In most systems of communication, this modulating waveform must be transmitted and received at high signal-to-noise ratio in order that the information be recovered with a high degree of accuracy. For example, in systems using amplitude modulation, recovery of the modulating waveform at high signal-to-noise ratio requires a high signal-to-noise ratio in the transmitted bandwidth also.

The types of systems which will be considered are those in which the modulation signal-to-noise ratio may be assumed to be sufficiently high at the output of the multiplex decoder for recovery of the messages, even though the signal-to-noise ratio in the transmitted time frequency region is low; and this must remain true even when several asynchronous channels are operating in the network.

Asynchronous multiplexing systems, then, are of interest insofar as they are capable of solving this communication problem efficiently, while at the same time providing certain other advantages. Two facts indicate that interesting possibilities do exist. One is the fact that coding schemes exist for error correcting and other forms of interference rejection, which may be viewed as restoring orthogonality at the output of a nonorthogonal system. The second is that many communication networks of the type diagrammed in Fig. 1 are such that very few of the stations of the network transmit during any given interval of time. Thus we may be able to design systems which will provide random access to the communication medium for any assigned station at any time, and maintain efficient communication with tolerable interference only if less than a certain number of the stations will be in use at any one instant. The interference becomes intolerable when the number of stations in use exceeds this critical number.

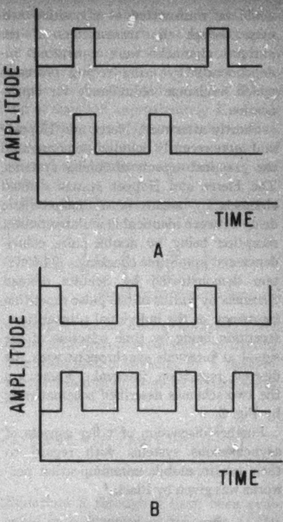


Fig. 3. Nonorthogonal multiplexing systems. Demultiplexing by

A—Spacing of repeated sample points (Pierce & Hopper[3])
B—Pulse carrier repetition rate

The purpose of this report is to summarize present knowledge of asynchronous multiplexing systems, the methods of analyzing them, to evaluate that knowledge, and to point out the most important problem areas remaining.

Survey of Early Literature

In 1950, White[1] gave a theoretical discussion of the transmission of information by a group of nonsynchronous pulse-type communication channels. His discussion was essentially a feasibility study of asynchronous multiplexing systems. He showed that some such systems do exist conceptually at least; that information

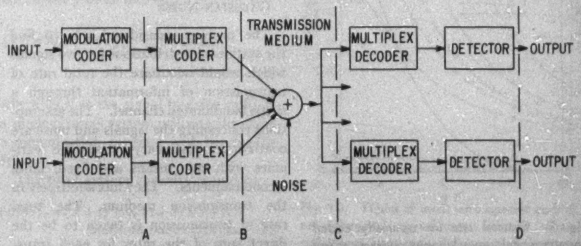


Fig. 4. General asynchronous multiplexing system

could be transmitted at a positive rate even though the transmitters of the various channels were operating independently by using coding processes which exchange redundancy for equivocation.

Shortly afterward, Pierce and Hopper[2] and more recently Schultz[3] demonstrated the practical aspects of similar systems. The Pierce and Hopper system defined channels by means of a characteristic delay between identical modulated pulses, reception being by double-pulse coincidence and amplitude checking. The system demonstrated by Schultz defined channels by means of the pulse repetition frequency of the individual transmitters, reception being by time selection of the signal at intervals synchronous with the desired repetition interval. These are the two schemes described schematically by Fig. 3.

Further discussion of other aspects of asynchronous systems with regard to their use in mobile communication networks was given by Black.[4]

Description and Evaluation of Theoretical Approaches

INFORMATION THEORY APPROACHES TO A STUDY OF ASYNCHRONOUS MULTIPLEXING SYSTEMS

There are two general areas in which the methods of information theory[5] apply to a study of asynchronous multiplexing systems. One of these concerns the synthesis of the best asynchronous system. It requires specification of the channel in terms of the statistics of the noise and certain other properties of the signals, such as their power limitations. Calculus of variations and other suitable methods of analysis are used to find the statistical distributions of the signal variables which provide optimum performance and a measure of this optimum performance in terms of rate of transmission. The details of a specific coding procedure, however, would not be given by such a study.

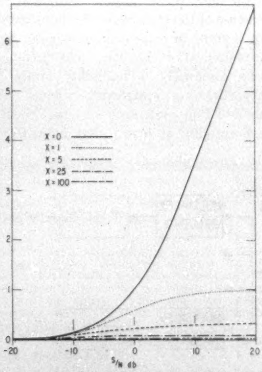


Fig. 5. Channel rate for optimally coded nonorthogonal asynchronous system versus S/N ratio. X+1 = number of channels in use

The other area concerns the use of information theory methods in the analysis of specific systems. Here, knowledge of the statistical properties of the actual signals and the noise or interference is required. The capacity of the system is then calculated in terms of source entropy and equivocation. Such a study will give the theoretical limitations of any similar system using codes which have the same statistical properties.

CODING THEORY APPROACH

A coding theory analysis can be made only if certain assumptions are made. A specific coding scheme must be chosen for which the coding theory is known, and the necessary properties of the communication system, in terms of the signals, noise, and interference must be such that the coding theory is applicable.

The coding theory study can then provide either an analysis of a specific system, or when carried out in terms of variable parameters, a synthesis procedure for the best system. It will also provide answers to the access questions; that is, how many stations can be assigned to the network, and how many can use the system simultaneously. The complexity of practical systems can be estimated in terms of what equipment will be necessary to perform the coding. The coding theory analysis will apply to practical systems only insofar as the assumptions do.

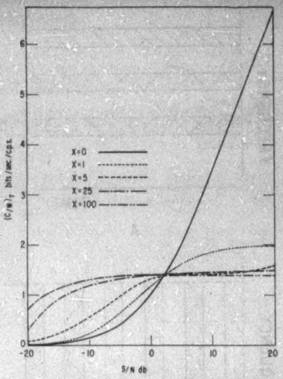


Fig. 6. Total system rate for optimally coded nonorthogonal asynchronous system versus S/N ratio. X+1 = number of channels in use

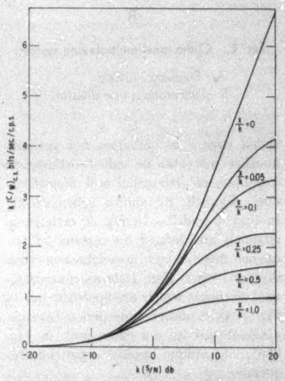


Fig. 7. Channel rate for optimally coded, nonorthogonal, channel-synchronized asynchronous system versus (KS/N) ratio. X+1 = number of channels in use. 1/K = fraction of channel

Applications of Information Theory

THE OPTIMUM STATISTICS FOR TWO CONTINUOUS SIGNALS IN ADDITIVE GAUSSIAN NOISE

The problem considered was to find the statistical distributions for two signals which would maximize the total rate of transmission of information through a single bandlimited channel. The assumptions concerning the signals and noise are continuous, stationary, stochastic variables with zero means and fixed, finite second moments. They mix additively in the transmission medium. The total rate of transmission is taken to be the direct sum of the rates for each transmitter, where one transmitted variable, say x, encounters a "noise" equal to the sum of the actual noise, n, and the other transmitted variable, y.

Maximization of the total rate is accomplished by applying calculus of variations methods. The condition necessary for a stationary rate is given by the solution of the Euler equations which are found to be nonlinear integral equations for which no general solution is known.

By direct substitution one can show that Gaussian distributions for the signals solve the Euler equations, giving a stationary rate equal to

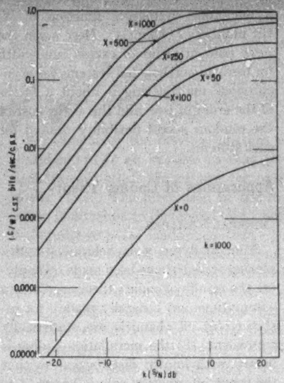


Fig. 8. Total system rate versus K(S/N) ratio for optimally coded nonorthogonal, channel-synchonized asynchronous system. X+1 = number of channels in use. 1/K = fraction of system degrees of freedom used by one channel

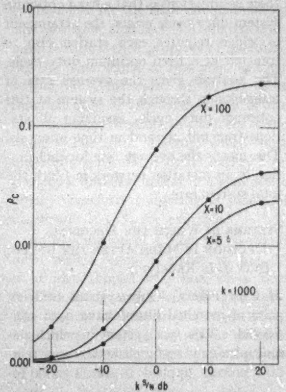


Fig. 9. Figure of merit ratio, channel synchronized versus optimal coded, nonorthogonal asynchronous system. X+1 = number of channels in use. 1/K = fraction of system degrees of freedom used by one channel

$$R/W=\log_2\left(1+\frac{S_x}{S_y+N}\right)\left(1+\frac{S_y}{S_x+N}\right)$$
bits/sec/cps (cycles per second) (1)

where S_x, S_y, and N are the second moments of the distributions of the variables x, y, and n, and W is the bandwidth of the channel. It has not been possible so far to determine the nature of this stationary rate, whether it is a relative maximum or minimum, or the sufficient conditions for it to be a maximum.

Bounds on the Performance of Asynchronous Systems

R. L. Shuey of the General Electric Company has made a study of the maximum rate at which information can be transmitted through a bandlimited transmission medium under the assumption that the coding is such as to make the interference in one channel a result of the signals of all other channels look like white bandlimited Gaussian noise. It might be expected that this requirement would be met easily when a large number of channels are in use when the central limit theorem can be made to apply. However, it is of more interest to consider the consequence of the assumption in the absence of that limit. The consequences are that each signal must be a Gaussian distributed variable[6] and the Shannon-Tuller law[5] must hold. This study avoids the question of the validity of the Shannon-Tuller law in the absence of synchronization. Its results are, however, probably valid bounds for infinite length messages.

Two further assumptions were made which make a more or less natural division between hypothetical coding systems for the purposes of analysis. One is that the various channels of the system are all sensitive to the same general regions of frequency time space, and the other is that each channel is sensitive to only a fraction of the total region of frequency time space accessible to and utilized by the system. The first assumption describes what will be called general asynchronous systems, while the second describes systems which will be termed channel-synchronized asynchronous systems.

General asynchronous systems, then, define channels in such a manner that each channel accepts as interference all the signal energy from all others. For a system of $X+1$ identical, independent channels each using signals of power S, the capacity of an individual channel is

$$C/W=\log_2\left(1+\frac{S}{XS+N}\right)\text{ bits/unit bandwidth/sec}\quad(2)$$

where S/N is the signal-to-noise ratio of a single channel system utilizing the same communication medium and bandwidth W.

The total system capacity would be

$$(C/W)_T=(X+1)\log_2(1+S/XS+N)\text{ bits/unit bandwidth/sec}\quad(3)$$

The channel-synchronized asynchronous system, on the other hand, defines channels in a manner such that each occupies, for example, only $1/k$th of the total system degrees of freedom. A channel whose signals are bandlimited to about W cps and of duration approximately T seconds has $2TW$ degrees of freedom or dimensions. The systems of Fig. 3 are examples of channel-synchronized asynchronous systems.

Thus the channel capacity for such a system is

$$(C/W)_{\text{C.S.}}=\frac{1}{k}\log_2\left(1+\frac{kS}{XS+N}\right)\text{ bits/unit bandwidth/sec}\quad(4)$$

since each channel also accepts only $1/k$th of the total system noise and interference. The total system capacity for a channel-synchronized asynchronous system would be

$$(C/W)_{\text{C.S.T.}}=\frac{X+1}{k}\log_2\left(1+\frac{kS}{XS+N}\right)\text{ bits/unit bandwidth/sec}\quad(5)$$

Equations 2 through 5 have been evaluated as a function of the ratio (S/N) and are shown in Figs. 5 through 8. The comparison of these two systems is of interest, and the ratio of total capacities can be written

$$\rho_c=\frac{\left(\frac{C}{W}\right)_{\text{C.S.T.}}}{\left(\frac{C}{W}\right)_T}=\frac{k}{1}\,\frac{\log\left(1+\frac{kS}{XS+N}\right)}{\log\left(1+\frac{S}{XS+N}\right)}\quad(6)$$

$\rho_c<1$ for all X, $k>1$, and is shown in Fig. 9.

Thus ρ_c is a measure of the price paid

GENERAL ASYNCHRONOUS SYSTEM VS. FREQUENCY DIVISION MULTIPLEXING SYSTEM

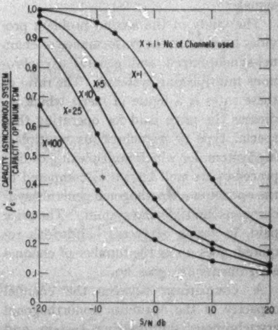


Fig. 10. Figure of merit ratio optimal coded, nonorthogonal asynchronous system versus orthogonal frequency division multiplex system. X+1 = number of channels in use

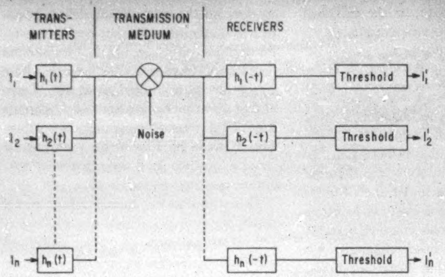


Fig. 11. Matched filter asynchronous multiplexing system

for the use of the channel-synchronized asynchronous system, and can be considered as a figure of merit for them. Systems of this type would, in general, be interesting when the parameters have values in the regions $S/N>10$, $X>10$, $(X/k)(S/N)<1$, say and then ρ_c is of the order of one tenth. The advantage of such a system is that it provides asynchronous channels in which the modulation is recovered at a high signal-to-noise ratio by using known coding methods. The disadvantages of the system lie in the implied requirement of excess bandwidth for transmission at reasonable rates.

Summary Remarks

In the study of the first problem, the methods of analysis lead to unsolved integral equations. The rate can be evaluated, however, in the closed form for the special case of Gaussian distributions, but the question arises: With respect to what statistical distributions is this a maximum? Does an optimum set of statistics exist for any general class of signals?

The study of the second problem provides bounds on the performance of channel-synchronized and general asynchronous multiplexing systems. The ratio of these rates is a figure of merit which indicates the price paid for operating that special type of asynchronous multiplexing system, for which multiplexing coding processes are well known as compared to the optimum nonorthogonal general asynchronous multiplexing system. The price paid, in terms of excess bandwidth required, goes up as the number of channel assignments (k) goes up.

A comparison between the channel capacity of the optimum nonorthogonal asynchronous multiplexing system and that for an optimally coded orthogonal system, utilizing frequency division multiplex, would indicate the price paid for nonorthogonal transmission as well. Thus we compare equation 3 to the channel capacity per unit bandwidth of a system also having $(X+1)$ channels used in frequency division. For optimal coding, the Shannon-Tuller law would hold for the ith channel, whose capacity is

$$C_i=W_i\log_2\left(1+\frac{S_i}{N_oW_i}\right)\text{ bits/sec}$$

where for the comparison, $N_o=N/W$ =noise power per unit bandwidth in the system transmission bandwidth, $W_i=W/X+1$ =bandwidth of the ith channel, $S_i=S$ signal power in each channel. Then the total system capacity when $X+1$ channels of the frequency division multiplexing (FDM) system are used is

$$C_T=W\log_2\left(1+\frac{(X+1)S}{N}\right)\text{ bits/sec}$$

and

$$\frac{C_i}{W_i}=\frac{C_T}{W}=\log_2\left(1+\frac{S(X+1)}{N}\right).\text{ bits/sec/cps}$$

The ratio of capacity of the optimum nonorthogonal asynchronous system to that of the optimum (optimally coded) FDM system is

$$\rho_c'=\frac{(X+1)\log\left(1+\dfrac{S}{XS+N}\right)}{\log\left(1+\dfrac{(X+1)S}{N}\right)} \tag{7}$$

Equation 7 is shown plotted in Fig. 10. A typical range of interest is where S/N is large, say $(S/N)=10$, and with X in the range 10 to 1,000.

Then

$$\rho_c'\rightarrow\frac{1}{4.5}\text{ to }\frac{1}{10}.$$

$$\text{For very large } X,\ \rho_c'\rightarrow\frac{X\log_e\left(1+\dfrac{1}{X}\right)}{\log_e\left(1+\dfrac{XS}{N}\right)}\rightarrow\frac{1}{\log_e\left(1+\dfrac{XS}{N}\right)}$$

a monotone decreasing function of X, the number of channels. Here again the price paid in terms of excess bandwidth for using nonorthogonal systems for asynchronous multiplexing is about a factor of the order of five, and this is the cost of the random access flexibility desired for such systems.

Applications of Coding Theory

General Remarks

Some analyses of asynchronous multiplexing systems have been made by applying the results of coding theory. Certain assumptions are generally made; for instance, (a) all channels are statistically equivalent, (b) the modulation coding is binary with known distribution so that the input information rate is known. It has further been assumed that the multiplex coding involves the restriction to fixed length (time sequences) and/or fixed energy signals. The multiplex coded signals may involve alphabets of two symbols, three symbols, or they may be assumed continuous. A consequence of these assumptions is that a fixed optimum system duty cycle exists, the attainment of which requires each station also to transmit at a fixed optimum duty cycle. The analysis gives the average rate of transmission through the system at this optimum duty cycle, assuming all stations transmit at random time intervals. This means the average rate for each station in an n station network is $1/n$th the total system rate.

Systems in Which the Decoding Decision Is Made After the Entire Signal Is Received

Matched Filters. Only systems utilizing pairs of matched filters[7] have been considered. This is a system in which demultiplexing is accomplished by a correlation process applied to the signals re-

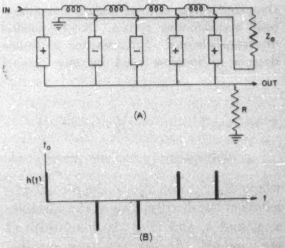


Fig. 12. (A)—Idealized matched filter. (B)—Pulse response

ceived. Fig. 11 shows schematically such a system. The filter used at the transmitter codes binary information into a signal $h(t)$ which is a function of its impulse response and the form of the driving function which is the binary alphabet symbol 1.

For simplicity the driving function is considered to be an impulse so that the signal output $h(t)$ is the impulse response of the filter. The filter used at the receiver is matched to this signal by having its impulse response made equal to $h(-t)$, the time reversal of the desired signal. This use of the matched filter as the stored reference for the correlation process has the advantage that no time synchronization between the receiver and the transmitter is required for the process of correlation to take place, since the filter correlates the received signal continually with the stored reference.

A matched filter performs the optimum operation on a mixture of known signal and additive Gaussian noise in the sense of extracting a maximum amount of information concerning the signal while destroying the irrelevant information due to the noise, for the class of equal energy equi-probable signals.[7,8] A related optimum feature is their property of maximizing the peak signal to rms noise in their output.[9,10] This ratio of (S/N) is a function only of the total energy of the matched signal and not of the form of the signal. The form of signal which most uniformly spreads the signal energy over the signal duration is that of a symmetric binary waveform consisting of pulses of amplitude $\pm\sqrt{S}$, for example.

An idealization of a filter which would form such a signal is shown in Fig. 12 as an ideal tapped delay line, whose taps feed isolating amplifiers of gain ±1. The outputs of these amplifiers are added together to form $h(t)$ as shown which is mathematically described by the sequence of binary digits, 1, −1, −1, 1, 1.

What relationship such a multiplexing arrangement has to an optimum for any noise or interference distribution which is different from Gaussian is not known, but in view of the asymptotic approach of the distribution of a sum of a large number of variables to a normal distribution, such systems have been considered. Such a system becomes more and more advantageous from the standpoint of the number of possible channels available as the length of the filter (sequence) increases. As a practical matter, however long delay lines imply distortion of the delay signal. The question arises as to how the use of nonlinear regenerative devices along the delay line would affect its operation. An analysis of the action of such a matched filter preceded by a 2-level quantizer has been made by Shuey and shows that such a quantization of the input to the filter, which is a sum of white Gaussian noise and a symmetric binary signal $(\pm\sqrt{S})$ does not affect the operation of the filter so far as the output signal-to-noise ratio is concerned to within a multiplicative constant. Specifically, the output signal-to-noise ratio is found to be of the form

$$K\left(\frac{2E}{N_o}\right)^{1/2}$$

with K of the order of 1/2 when $S/2WN_o$ is small enough that $[1\text{-}Erf(S/2WN_o)]$ can be approximated well by the first term in its expansion. Here S is the signal power, E the total energy in the signal, WN_o is the total noise power in the band used (of width W cps).

Uncorrelated Sequences. A study was made[11] of how many sequences of binary digits could be picked such that both their autocorrelation and their cross-correlation sequences were bounded in absolute magnitude with the length of the sequence N and the bounding value T as parameters. This is the question: How many channels can be assigned to a multiplexing system of the type shown in Fig. 11? The matched filters here are assumed to be linear, without quantizers so that with no input, no ouput is obtained. Thus the input alphabet to each receiving filter consists of the symbols 0, +1, −1. A binary input to the ith transmitting filter is a positive or negative impulse, causing $\pm h_i(t)$ to be transmitted, where $h_i(t)$ is a sequence of binaries ±1. A 0 is assumed to be the input at all other times.

All correlation sequences required for the analysis of this system, taking into account known symmetry properties of the correlation sequence, were calculated for $N=5$ and 6. An exhaustive selection process was then used to find maximal sets of binary sequences whose correlation sequences obeyed the criterion of being less than or equal to T in magnitude, except for the central term of every autocorrelation sequence. This requirement allows recovering synchronization information from the signal and is excessively severe for the simple communication problem. The results of the investigation are given in Fig. 13 where $M(N,T)$ is the maximum number of members in a set such that the correlation of all members behave as required. The monotone properties of $M(N,T)$ are evident. It was shown that $M(N,T=N-2)\approx 1+2^{N-1/2}$, and that in principle one can extend the analysis from a consideration of N digit sequences to $N+1$ digit sequences without a direct calculation of the correlation sequences. The analysis was then extended to 7-digit sequences, but was extended no farther for practical reasons.

Randomly Selected Binary Sequences. In order to solve this access question in a range of sequence length which is of more

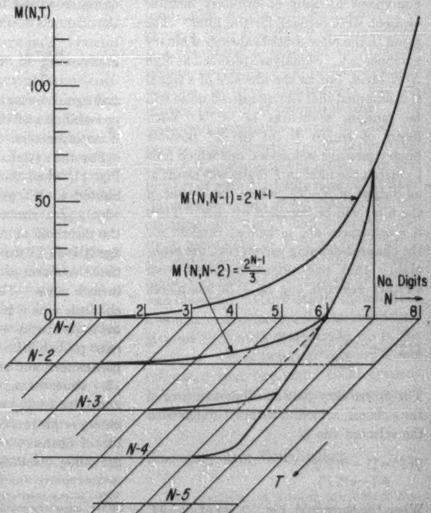


Fig. 13. Number of binary sequences M(N, T) in a set whose correlations are less than +T versus N the number of digits in the sequence

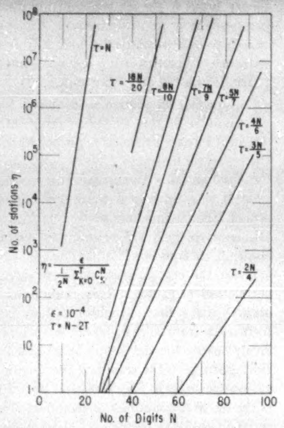


Fig. 14. Number of randomly selected binary sequences selectable without exceeding a probability ϵ of having a pair of them nonseparable with threshold parameter $\tau=(N-2T)$, $\epsilon=10^{-4}$

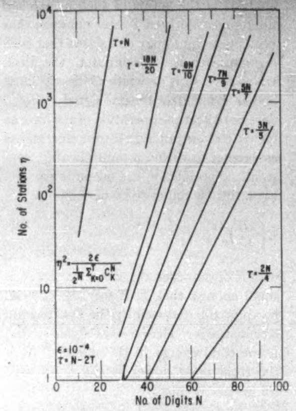


Fig. 15. Number of randomly selected binary sequences selectable without exceeding a probability 2ϵ of having any pair of the sequences be nonseparable with threshold parameter $\tau=N-2T$, $\epsilon=10^{-4}$

interest, Shuey's report considered the problem of how many binary sequences could be randomly selected and still cause interference with low probability. The specific model of a system considered, consisted of a 2-level quantizer preceding such a tapped delay line filter. It is clear that except for timing requirements, this is equivalent to using an ordinary bistable element shift register for the filter. The input to the filter consists always of binary symbols ± 1. Signals are received at high S/N ratio, but in the absence of a signal it is assumed that the quantized noise will be random sequences of ± 1's. With filters of length N, of the 2^N possible binary message sequences, one which fails to match the filter in T digits will result in an output of $N-2T$, where the output of the filter will be bounded by $\pm N$. Then if η sequences are picked at random for the channel-defining waveforms, the probability that one sequence differs from another selected one in T or less places is

$$P(T)=2^{-N}\sum_{k=0}^{T}C_k^N$$

The probability that no other sequence of the η chosen coincides within T digits with the selected one is

$$Q(T)=[1-P(T)]^{\eta-1}$$
$$\doteq 1-\eta P(T)$$

When the threshold, Fig. 10, set at $N-2T+\Delta$, at the output of the filter is exceeded, interference* takes place and the probability of this interference is

$$P(i)=1-Q(T)=\eta P(T)$$

Thus for a given probability of interference $P(i)=\epsilon$, η can be calculated with N and T as parameters.

The probability of interference in any of the channels because of any other of the channels would be given by

$$P_{tot}(i)=\frac{\eta(\eta-1)}{2}P(T)$$

and again, η can be calculated for a given probability ϵ of interference, with N and T as parameters.

For the system as described, shown in Fig. 11, when the random noise is bandlimited to W cps, then sequences occur which give rise to false alarms (exceed the threshold at $N-2T+\Delta$) on the average $2W\ P(T)$ times per second; $P(T)$ is the false alarm probability for the system in such noise. The maximum number of channels η as a function of the length of the sequences N were calculated with various probabilities of error ϵ and T as parameters, and are given in Figs. 14 and 15. Here we see that such a system of multiplex coding using randomly selected binary sequences can provide a large number of channels with low probability of providing insufficient channel separation if the length of the sequence is 50 to 100 digits.

This analyis is intended only to answer the access question by putting bounds on the number of channels which can be defined by this system of multiplex coding. Clearly the system could not utilize all channels simultaneously, but would operate with reasonably low equivocation only at a very low system duty cycle. Such practical questions have been ignored as to how actual interchannel interference caused by simultaneous reception of two of the binary waveforms with some degree of overlap would affect the rate of transmission. With a 2-level quantizer, preceding the filter, a stronger signal from one channel will swamp out a weaker one. In receiving the weaker one, on the average only half the overlapping digits will be in error, so if the filter and threshold detector can allow T errors, an overlap of $2T$ digits would be allowed on the average without causing interference. More overlap of signals would cause interference. If the filter is not preceded by a quantizer, a modified approach to the problem of rate of transmission would be required. In this case, the S/N ratio varies as $\sqrt{N}$ since interference is like a random walk problem.

SYSTEMS USING FILTERS MATCHED TO WHITE NOISE

The possibility of using very long samples of random noise for defining channels arises.[5,12] For a very long sample of such a waveform, e.g., a segment of white Gaussian bandlimited noise of duration T seconds, limited to a band of W cps, the autocorrelation is effectively of the form

$$\frac{\sin 2\pi W(t-n/2W)}{2\pi W(t-n/2W)}.$$

Such a function has only one nonzero sample value, and can be considered to be the ideal short-time autocorrelation function.[8] Such a waveform could be used to represent a binary symbol in one channel. The noise waveforms and their matched filters can be specified by their $2TW$ (random amplitude) sample values. The various waveforms would be of nearly equal energy, and interference would be additive Gaussian, so that such a system of matched filters would constitute a system of optimum detectors in the Woodward sense of maximizing inverse probability of the signal if each signal is transmitted with equal probability.[7] Although there is an infinite number of such waveforms, only about $N=2TW$ of them can be in the transmission medium simultaneously, for then the output signal-to-noise ratio of a matched filter is reduced to unity. There is nothing in this systems idealization which prevents a single waveform from being transmitted repetitively N times in the interval T, and being correctly received, so long as no other transmissions take place simultaneously and the proper reception can take place at the unity signal-to-noise ratio.

---

* This is not interference in the strict sense, but rather misinterpretation of the received signals, i.e., thinking station S_{tk} transmitted when it was S_{jl}.

Only in the limit of infinitely long noise sequences are the cross correlations always zero. The practical question should be considered of how fast the correlation functions approach these optimum forms as the lengths of the sequences grow. Specifically, the rate at which the trading of redundancy for equivocation can take place in such a system as the length of the waveform segments is increased should be evaluated. This would require at least probabilistic knowledge of the envelope of the autocorrelation functions in order that a proper evaluation of the probabilities of error can be made.

Systems in Which Decoding Is by Pulse-to-Pulse Computation

Several papers have appeared recently in which coding theory analyses of specific systems are presented. The following general assumptions are usually adopted: Channel-defining binary codes are selected. Interference rejection is based on Hamming distance concepts or some other pulse-to-pulse computational requirement. Such a code will be used to transmit binary-coded information. Decoding with small probability of error requires very high signal-to-noise ratio, so that very low system duty cycles are required. This in turn implies that only very low channel rates will be afforded by such systems. Every channel is assumed to be the same as any other, and the binary information is transmitted at random instants. P. M. Lewis[13] considered a quantized binary channel operating at high signal-to-noise ratio. The signals to interference ratio is set by the duty cycle of each channel d, the length of the channel defining codes L, and the number of possible and actually used channels p and m, respectively. Average system performance was calculated in terms of the expected number of errors for a given choice of parameters p, m, L, d. For specified p and m, L and d were varied to maximize the total rate of transmission. The analysis proceeds by assuming L digits of a code are sprinkled at random throughout n digit spaces, whence $d=L/n$. Part of the L digits are some that are added to attain a specified low probability of false alarm, P_f. Using Hamming distance concepts coupled to the expected number of errors, values of L and d required to maximize the system rate of transmission H are found. Lewis' results are summarized:

$$d_{\text{opt}}=\frac{1}{2m}$$

$$H=\frac{1}{2L}$$

$$n=2mL$$

$$L=4e\log_e 2p+\log_2\frac{1}{P_f}$$

Two ranges for system specifications are of interest, and optimum "expected value" performance of such a system may be summarized by the following table:

System Specification	$p=25$ $m=10$ $P_f=10^{-8}$	$p=5{,}000$ $m=10$ $P_f=10^{-8}$
L	70	130
H	0.0077 bits/digit space	0.0038
n	1,400	2,600

The disturbing feature of this system is the large value of n required and the small system rate of transmission as well as the much smaller channel rate of 1 bit/n digit spaces.

Many of the disadvantages of this system are due to the quantizer at the receiver input, required by the postulated system of error detection decoding which allows only two received symbols. The requirement that the L code elements be randomly distributed among the n digit spaces implies that there be n digit storage available at the coding and decoding devices.

Noble[14] analyzed binary-pulse systems from the point of view of information theory, calculating the probabilities necessary for obtaining Shannon's information rate. The distinction was made between "symbol information rate," which measures rate excluding the effects of message coding, and "message information rate," which is the information rate available with the associated message coding constraints. For the pulse systems, the optimum symbol rate is exactly that given by White,[1] 0.7 bit/pulse width, while the application of the message coding constraints (equal number of 1's and 0's in the message were assumed) cuts this down to 0.2 bit/pulse width. The message coding is clearly an error-detecting system which makes no attempt to correct errors, and the loss of information rate caused by possible errors is surprisingly large. Error-correcting codes can increase the information rate by minimizing such loss. For a system using matched-filter coding to represent the binary symbols, an optimum symbol rate was found to be about 0.5 bit/sec/unit bandwidth-equivalent to 0.25 bit/pulse width in a bandlimited pulse system.

Experimental Evaluation of a Channel Synchronized System

A channel-synchronized asynchronous system, suggested by R. L. Shuey, utilizing delta modulation,[15,16] was simulated by E. R. Schultz[9] as a 2-channel system. Delta modulation is a form of pulse code modulation in which information is carried by the presence or absence of a pulse, or equivalently by pulse polarity in a constant *prf* pulse train. The pulses represent the change in the modulating waveform between sampling instants. Either a single or a double integration of the pulse train can be used as an approximation to the information waveform in the coding or decoding process. The outstanding advantage of the delta modulation system is the simplicity of the decoding equipment, consisting of the integrating circuits. The disadvantage lies in the very high sampling rate required for good signal-to-noise ratio (about 100-kc *prf*'s were used by Schultz).[9] Each channel of the multiplex system was distinguished by its unique pulse-repetition frequency. The advantage of using such a system for multiplexing, as pointed out earlier, is that a synchronization problem which can be solved is substituted for a coding problem whose solution is not known.

The 2-channel experimental system used pulse widths and *prf*'s which give about a 10% probability of interference, and simulated a practical multiplexing system operating with perfect synchronization between transmitters and receivers, i.e., perfect demultiplexing. The noise and interference were of three expected characteristic types: heterodyne, cross talk, and quantizing noise. The first two types were markedly dependent upon the relative values of *prf* used by the two channels, and were found experimentally to be minimized by a 20-kc difference in *prf* assignments. At this point a maximum S/N ratio of about 25 db (decibels) was observed. Calculations showed that a similar system having six channels would give an S/N ratio of about 11 db in the channels. As a practical matter, it was estimated that one could build systems having about 10 channels for use simultaneously.

Summary and Conclusions

The question of finding statistical distributions for signals which will maximize the rate of transmission of information in an asynchronous multiplexing system has been considered. A study of the general nonorthogonal system which uses Gaussian noise like coding, and that of the channel-synchronized asynchronous systems shows that such systems are inefficient from the information rate point of view compared to orthogonal systems such as frequency division multiplex. Such studies attempt to answer the question: What is the optimum asynchronous system, and how can its performance be measured? From the standpoint of information capacity, the answer to these questions would give a yardstick by which any practical system could be measured.

Lacking a complete answer to the questions, one is forced to consider specific systems. The coding theory analysis of various fixed-length block-coding schemes have been made for two purposes; for measuring the optimum information rate, and for finding the number of channels made available by the coding scheme. The question of how well finite-length matched-filter coding systems solve the access problem requires further study of the interference produced by the overlapping of channel signals in time of transmission.

The analysis presented here of the interference effects in binary systems treats only the case of video pulses, and will be changed if other transmission methods are postulated, such as for example, a binary frequency shift system. For very long binary sequences, new practical considerations arise, e.g., concern for local oscillator stability, since alphabet symbols ±1 imply phase sensitive detectors when a common r-f band transmission is used. These and other practical considerations have not been treated here.

References

1. Theoretical Aspects of Asynchronous Multiplexing, **W. D. White.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 38, 1950, p. 270.

2. Nonsynchronous Time Division with Holding and with Random Sampling, **J. R. Pierce, A. L. Hopper.** *Ibid.*, vol. 40, 1950, p. 1079.

3. An Experimental Time Division Asynchronous Multiplexing System Using Delta Modulation Pulse Coding, **E. R. Schultz.** *Report no. 55-RL-1408*, General Electric Research Laboratory, Schenectady, N. Y., 1955.

4. Modulation Theory (book), **H. S. Black.** D. Van Nostrand Company, Inc., Princeton, N. J., 1953.

5. A Mathematical Theory of Communication, **C. E. Shannon.** *Bell System Technical Journal*, New York, N. Y., vol. 27, 1948, pp. 379–423, 623–56.

6. Mathematical Methods of Statistics, **H. Cramer.** Princeton University, Princeton, N. J., 1951, pp. 212–13.

7. Information Theory and Inverse Probability in Telecommunication, **P. M. Woodward, I. L. Davies.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. III, 1952, p. 37.

8. On the Properties of Matched Filters, **D. W. Lytle.** *Technical Report no. 17*, Electronics Laboratory, Stanford University, Stanford, Calif., June 1957.

9. Analysis of Factors which Determine Signal-Noise Discrimination in Pulsed Carrier Systems, **D. O. North.** *Report no. PTR-6C*, Radio Corporation of America, Princeton, N. J., June 25, 1943.

10. A Theoretical Comparison of the Visual Aural and Meter Response of Pulsed Signals in the Presence of Noise, **J. H. Van Vleck, D. Middleton.** *Journal of Applied Physics*, New York, N. Y., vol. 17, 1946, p. 940.

11. Multiplexing with Uncorrelated Pulse Sequential Waveforms, **J. E. Taylor.** *Report no. 56-RL-1590*, General Electric Research Laboratory, Aug. 1956.

12. Communication in the Presence of Noise, **C. E. Shannon.** *Proceedings*, Institute of Radio Engineers, vol. 37, 1949, p. 10.

13. Coding Theory Analysis of a Class of Asynchronous Multiplexing Systems in a Quantized Binary Channel, **P. M. Lewis.** *Report no. 57-RL-1841*, General Electric Research Laboratory, 1957.

14. Mathematical Analysis of the Non-Synchronous Multiplex System, **M. Noble,** *Technical Report no. R57ELS65*, General Electric Company, 1957.

15. Delta Modulation, A Method of PCM Transmission Using the One Unit Code, **F. de Jager.** *Philips Research Reports*, Eindhoven, Netherlands, vol. 7, 1952, p. 442.

16. Delta Modulation, A New System for Telecommunication, **F. G. Schouten, F. de Jager, J. A. Greefekes.** *Philips Technical Review*, Eindhoven, Netherlands, vol. 9, 1952, p. 237.

Vol.
Pt. 1
1959

Ultrasonic Frequency Power Generator Using Transistors

YIN-MIN WEI
NONMEMBER AIEE

INCREASED USE of ultrasonic energy in industrial applications has created a greater demand for ultrasonic frequency power supplies. Most of these power supplies are constructed with vacuum tubes as the conversion devices. However, power supplies, which are made with transistors operated as high-current switches,[1] can be made higher in power conversion efficiency and more compact for the same power rating. Described in this paper will be the circuits which have been utilized in the generation of ultrasonic frequency power.

Methods of Generation

Two-Transistor Oscillator

The first circuit to be considered will be a 2-transistor push-pull oscillator utilizing a series tuned feedback circuit as shown in Fig. 1. This circuit which uses switching transistors was suggested by R. P. Putkovich as a means of providing a stable frequency oscillator. The operation of this circuit may be explained in the following manner. Assume transistor Q_1 is conducting. The voltage across power transformer T_1 primary equals $2E_B$. This same voltage appears across the feedback circuit of L_1, C_1 and feedback transformer T_2 primary (see Fig. 2). Initially the inductor L supports almost all the voltage (assume the feedback circuit has high circuit Q). As the inductor current increases, the voltage across T_2 primary rises. Feedback transformer T_2 secondary voltages are such as to keep transistor Q_1 conducting and transistor Q_2 cut off. As the condenser C_1 charges, the voltage across T_2 drops. T_2 secondary voltage drop causes Q_1 to start swiching OFF. The capacitor then begins to discharge through primary of T_1 and the voltage across T_2 is reversed. This initial reversal of voltage causes Q_2 to conduct and Q_1 to cut off, and the previous half cycle is repeated for Q_2 conducting. Frequency variation can be controlled by varying L_1 or C_1 or both. This oscillator can be used to drive a low-power load directly or to drive another higher powered inverter.

Four-Transistor Bridge

Another type of power inverter consists of a 4-transistor bridge circuit[2] (Fig. 3) which is driven by some lower powered oscillator. Drive signals appear at four separate, but identical, secondary windings (designated *a-b-c-d*) of drive transformers T_3 with polarities selected such that transistors Q_3 and Q_5 are driven into

---

Paper 59-1203, recommended by the AIEE Semiconductor Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted June 16, 1959; made available for printing August 19, 1959.

Yin-Min Wei, formerly with Westinghouse Electric Corporation, Pittsburgh, Pa., now resides in Madison, Wis.

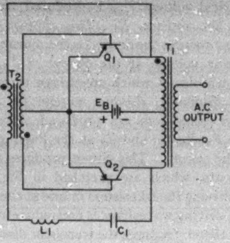


Fig. 1. Push-pull oscillator

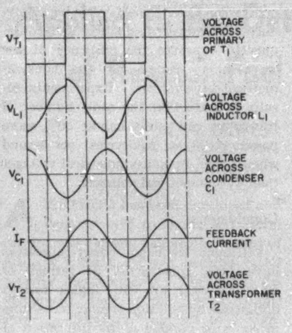


Fig. 2. Voltage and current waveforms in Fig. 1

conduction while Q_4 and Q_6 are driven to block. When transistors Q_3 and Q_5 are conducting, the input direct current flows from A to B in the primary of output transformer T_4. When transistors Q_4 and Q_6 are conducting, the input direct current flows from B to A in the primary of T_4. Thus, alternating voltage at the drive frequency is available at the secondary of output transformer T_4. Because each nonconducting transistor has to block the input direct voltage, the magnitude of the input direct voltage is limited by the maximum V_{Ce} rating of the transistor.

SERIES-CONNECTED BRIDGES

Because of the difference in switching speed, each of the transistors which are connected in series to operate from a direct voltage may take a peak voltage equal to the full supply voltage. Thus, series operation of transistors is not generally feasible. In order to accommodate higher input direct voltages, two or more transistor bridge circuits can be connected in series across the direct voltage input as shown in Fig. 4. The four transistors numbered *3*, *5*, *7*, and *9* are driven ON simultaneously by signals fom T_3, while transistors numbered *4*, *6*, *8*, and *10* are driven OFF simultaneously by opposing signals from T_3. When the signals from T_3 change polarity, transistors *3*, *5*, *7*, and *9* are driven OFF while transistors *4*, *6*, *8*, and *10* are driven ON. Thus, an alternating square-wave voltage is obtained at the output transformer T_4. Since the voltage across a condenser cannot change instantaneously, the condensers C_2 and C_3 help to equalize the direct voltage across the bridges in case the transistors in one of the bridges should turn ON or OFF sooner than the other because of differences in transistor switching times. Again, with the addition of the diode bridge at the center of each transistor bridge, the circuit would continue to work even if one of the transistors is disconnected from the circuit. Assume Q_3 is taken out of the circuit. When the odd numbered transistors are driven ON only the lower bridge works and supplies the output power from condenser C_3 through the primary winding P_2 of transformer T_4. This makes the voltage of C_3 lower than that of C_2. But when the drive voltage changes polarity, only the even numbered transistors are conducting and both the transistor bridges work. This time, the voltage from C_2 through the upper bridge to winding P_1 of T_4 is higher than that from C_3 to P_2 through lower bridge. Since P_1 and P_2 are identical windings on the same transformer and have the same polarity, C_2 actually supplies the output power and charges C_3 through P_1 and P_2 and diodes D_6 and D_8 to equalize the voltage of the two condensers It is understood that this same action through the primary windings of the output transformer also tends to maintain equal voltage across C_2 and C_3 when the voltage difference is caused by some other factor, such as the d-c leakage of the two bridges In addition, the diode bridges provide a path for the continuous flow of current carried by an inductive load such as an ultrasonic transducer, when the transistors are turned OFF.

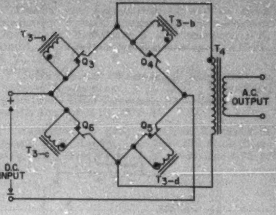


Fig. 3. Four-transistor bridge

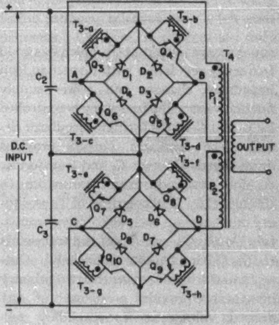


Fig. 4. Series-connected bridges

System for Power Generation

Combining the oscillator of Fig. 1 with the bridge-type power inverter of Fig. 4, a system for the generation of a-c power was obtained and is shown in Fig. 5. The frequency control oscillator provides a

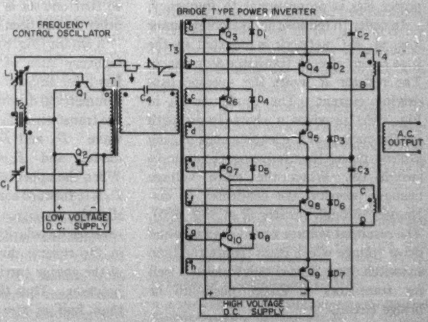


Fig. 5. System for power generation

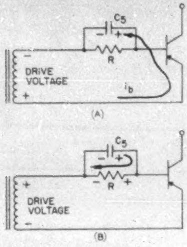


Fig. 6. Base biasing circuit

square-wave output at T_1 which has good frequency regulation with respect to load impedance changes and power supply fluctuations. The frequency is controlled by L_1 and C_1 in the oscillator feedback circuit. The oscillator output provides drive signals through C_4 and transformer T_3 for power inverter transistors Q_3 through Q_{10}. The polarities of T_3 secondary windings are such as to drive transistors Q_3, Q_5, Q_7, and Q_9 ON and transistors Q_4, Q_6, Q_8, and Q_{10} OFF alternately. Output transformer T_4 sums the two primary inputs and provides a-c power at any desired voltage at its secondary terminals. Good regulation and filtering of the high-voltage power supply is required for reliable operation of the power inverter due to the voltage limitations of the power transistors.

Circuit Improvements

R-C Network in Base

The transistors in the system described are operated in the switching mode to minimize power dissipation in each transistor. As the operating frequency increases, the power loss during switching becomes a substantial part of the total power dissipation because switching power loss is proportional to frequency.

In order to reduce this switching power loss, an *R-C* network is inserted in the base circuit of each transistor as in Fig. 6. The resistor R limits the average base driving current. The condenser C_5 in Fig. 6(A) provides a short-circuit path for the current at the leading edge of each driving pulse, thus reducing the time to switch ON. When the drive voltage changes polarity, the condenser C_5 discharges through resistor R of Fig. 6(B). The resultant voltage across R adds to the drive voltage which helps reduce the time to switch OFF. This circuit functions well for transistors in either push-pull or bridge circuits.

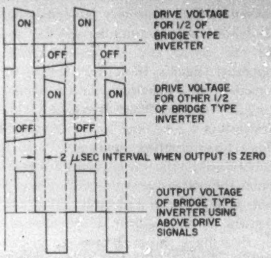


Fig. 7. Nonsymmetrical drive waveforms

Capacitance Drive Circuit

It has been shown by Moll[3] that for equal driving current transistors are switched ON faster than OFF because of the minority charge carriers stored in the base region. The *R-C* network in the base circuit helps to reduce switching time both ON and OFF. However, a condenser C_4 in series with the driving transformer as shown in Fig. 5, produces an exponentially decaying driving voltage, thus minimizing the base charge carrier storage to reduce the switching OFF time. This, in turn, reduces the conduction time of all transistors and helps reduce the instantaneous short circuit which appears across the d-c supply twice each cycle.

Nonsymmetrical Driving

In a push-pull or bridge circuit, the d-c power supply will be instantaneously short-circuited twice each cycle when all the transistors are in the ON state because they are switched on faster than off. This periodical, temporary short circuit produces high-power dissipation in the transistors. To guard against this short-circuit situation, the two sets of transistors, which are operated ON and OFF alternately, may be driven separately by nonsymmetrical voltage waveforms (Fig. 7) so that one set is switched off before the other set has been switched on.

A circuit (Fig. 8) which makes use of the recovery time of semiconductor diodes[4] has been employed to generate the nonsymmetrical driving signal. The input to the transformer T_5 is a symmetrical square wave. D_9 and D_{10} are identical silicon diodes and $R_1 = R_2$, $R_3 = R_4$, $C_6 = C_7$. When the upper terminal of T_5 is positive, D_9 is in forward conduction and V_{AB} is only slightly positive. As the input voltage reverses its polarity the diode D_9 conducts in the reverse direction intially because of the charge carrier storage in the diode junction. Thus the diode conducts more than half of the time and a nonsymmetrical voltage V_{AB} is produced across the the resistor R_3. This voltage, in turn, produces the desired nonsymmetrical driving voltage at the output of T_6 for the transistors which are driven through C_6. Similarly, driving voltage for the other transistors are produced through D_{10}, R_2, R_4, C_7, and T_7 at the lower part of the circuit. This circuit produces two outputs, which are sketched in Fig. 7, for driving the transistors in bridge circuit. The driving waveshape is nonsymmetrical and thereby reduces the transistor dissipation during switching.

The addition of diodes D_{11} and D_{12}, as indicated by the dotted lines to shunt R_1 or R_2 when D_9 or D_{10} is reverse biased, reduces the power dissipation and consequently improves the efficiency of the waveshaping circuit for the nonsymmetrical drive voltage. Experimental results indicate that the same number of transistors can safely handle 50% more power at 20 kc (kilocycles) per second when driven by nonsymmetrical voltages than by symmetrical voltages.

Conclusions

Ultrasonic frequency power generators have been constructed at ratings of 50 watts to 1 kw with higher ratings obtainable by addition of more transistors in series or parallel or by utilization of transistors having higher voltage and current ratings.

The frequency of these germanium transistor units is a function of temperature, supply voltage, and load. The frequency stability obtained from transistorized power generators is shown as:

1. 1% frequency change for 10 F (degrees Fahrenheit) temperature change.
2. 1% frequency change for 10% supply-voltage change.
3. 0.02% frequency change from no load to full load.

A 1-kw unit having dimensions of 20 inches width, 9 inches height, and 15 inches deep has been constructed. It weighs less than 50 pounds and has a d-c

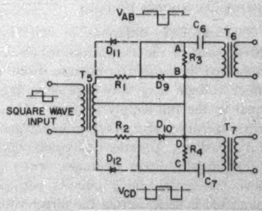


Fig. 8. Nonsymmetrical drive voltage shaping circuit

Vol.
Pt. 1
1959

to a-c power conversion efficiency of about 85%. A generator of the same power rating using vacuum tubes may be three times its size and weight and only about one-half as efficient. At the present time, the cost of transistor units is higher than vacuum tube units; however, this situation is expected to improve when higher power transistors become available. As a matter of fact, laboratory sample of silicon transistors rated at 300 volts and 20 amperes are made by Westinghouse Electric Corporation.

References

1. Junction Transistors Used as Switches, R L. Bright. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, Mar. 1955, pp. 111–21.

2. Design and Application of a New High Frequency Power Source for Fluorescent Lighting, W. H. Johnson, J. L. Winpisinger, J. F. Rossel, Jr. *Illuminating Engineering*, New York, N. Y., 1958.

3. Large Signal Transient Response of Junction Transistors, J. L. Moll. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 42, Dec. 1954, pp. 1773–84.

4. Handbook of Semiconductor Electronics (book), L. P. Hunter. McGraw-Hill Book Company, Inc., New York, N. Y., 1956, pp. 4-20 to 4-23.

Delayed Neutron Detection Methods Applied to the Detection and Location of Reactor Fuel Element Failures

R. A. DEWES
NONMEMBER AIEE

J. C. CHILDS
NONMEMBER AIEE

A METHOD has been developed and demonstrated to detect and locate reactor fuel element failures by means of primary coolant delayed neutron (DN) activity measurements. The initial development, done at the Knolls Atomic Power Laboratory,* applied to a high-temperature, pressurized, recirculating water-cooled reactor system. However, it can be applied to a variety of other types of reactor systems, including air-cooled types. By this method, the state of reactor fuel integrity can be continuously and simply monitored. The actual location of fuel element failures can also be determined; in some instances this location can be accomplished without the necessity of shutting down the reactor.

In brief, the method consists of measuring the neutron activity of a flowing sample of primary coolant fluid as it passes through a DN detection station. The presence of neutron activity above a certain predetermined background level indicates the presence of fission product emitting surfaces in the reactor system. This indicates the presence of a failed fuel element. The neutron activity arises from the DN emitting fission product isotopes I-137 and Br-87 which enter the coolant stream by recoil from an exposed fuel surface.

In order to locate the failure within a multicellular reactor core, either of two methods may be employed. The first of these involves auxiliary equipment which permits continuous flow of primary coolant samples from selected portions of the reactor core through individual sampling lines. A measurement of the neutron activity of the effluent from each of these lines indicates the location of the failure within the resolution of the sample line intake location. A second, less complicated, method has been found to be usable in this application and may be applicable to most reactors. This method, termed the "flux depression method," utilizes the local reduction in flux which accompanies the insertion of one or more control rods at constant average reactor power. Such a rod manipulation depresses the specific fissioning rate in the region of the inserted rod(s), with the spatial extent and amount of flux depression depending upon many parameters. If a failure is present in the region of the depressed flux zone, the average neutron activity of the primary coolant as measured by the permanently installed delayed neutron detecting station is found to decrease. The failure is thus located within the resolution of the flux depression pattern of various rod movements.

Discussion

Failure Detection

In a reactor system, fission product isotopes may enter the primary coolant system from fuel surfaces that are exposed to the coolant because of a failure, from fissionable atoms in or on the reactor construction materials or fuel cladding surfaces, or from fissionable atoms in solution in the primary coolant. Of the many fission products released, only a few emit neutrons. Of these, it is necessary only to consider the DN emitters given in the following:

Isotope	Fission Yield	Neutron Transition Fraction	Half Life
Br-87	0.027	0.02	56 sec.
I-137	0.049	0.06	22 sec.

These isotopes are the most important because of half-life. The other DN emitters have half-lives on the order of 5 seconds or less, and so have decayed to the point that they are not detectable by the time they have arrived at the detecting station mentioned in the previous sections.

There are various mechanisms associated with the release of fission product isotopes: gaseous diffusion, erosion, corrosion, and fission recoil. Of these, only the recoil mechanism is important in establishing the level of DN activity in the coolant of a reactor which has high integrity fuel. Fissionable atoms in solution may also contribute neutron activity; however, the contribution is negligible in a reactor primary system which has a normally efficient purification system and low corrosion and erosion rates. Thus, measurement of the neutron activity of a primary coolant sample is directly related to the magnitude of exposed fuel surfaces. A certain base level of neutron activity exists because of the presence of fissionable atoms on or near the surfaces of the core construction materials.

The use of the bromine and iodine DN emitting isotopes as a basis for this measurement provides many advantages over other methods of failure detection. Their

---

* Operated for the United States Atomic Energy Commission by the General Electric Company under contract W-31-109 Eng-52.

Paper 59-1184, recommended by the AIEE Nucleonic and Radiation Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted June 15, 1959; made available for printing August 31, 1959.

R. A. Dewes and J. C. Childs are with the General Electric Company, Schenectady, N. Y.

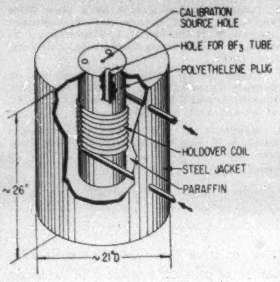


Fig. 1. DN detection station

short half-life means that an up-to-date picture of reactor conditions is available at all times, and laborious corrections for past reactor history need not be undertaken. Further, the neutron level is an unambiguous indication of fission product activity; there is no confusion with competing nonfission-product coolant activities that exist in, say, gamma analysis. Also, no primary coolant sample need be removed from the primary system nor handled in any manner; only the activity of a small leg with a holdover vessel need be measured.

The mathematical relation which governs the fission product release is fairly simple to evaluate. The activity per unit coolant volume of any isotope released is

$$B=\frac{ryp\sigma S\phi f\rho hA}{VM}\left[\frac{\lambda}{\lambda+d+\beta+l}\right](1-e^{-\lambda t}) \quad (1)$$

where

r = fission product isotope release factor = nuclei entering stream/nuclei produced ~0.25 for recoil mechanism
y = fission yield
p = neutron transition fraction
σ = U-235 fission cross section at reactor temperature
S = area of failure (exposed fuel)
ϕ = thermal neutron flux density (average)
f = fraction by weight of U-235 present in meat
ρ = fuel meat density
h = fission product recoil range in fuel meat
A = Avagadro's number.
V = active system volume
M = gram molecular weight of U-235
λ = decay constant for isotope
d = deposition removal constant for isotope
β = system purification constant
l = system leakage constant

In the case of the halogen DN emitters, the values of l, β, and d are negligible compared to λ. Also, at equilibrium $e^{-\lambda t}\sim 0$; thus

$$B=\frac{rypo\sigma S\phi f\rho hA}{VM} \quad (2)$$

The count rate (CR) observed at the detection station to which the sample takes T seconds to travel, having a volume v, and a detection efficiency ϵ, is

$$\mathrm{CR}=(Bv\epsilon)e^{-\lambda T} \quad (3)$$

The background in this detecting system is caused by environmental neutrons at the detecting station and by the previously mentioned DN's produced by the recoil nuclei from natural uranium impurities in the reactor core surfaces (of the order of a few parts per million unless contaminated). The only other neutron activity present is that caused by N^{17}; this isotope is produced by the $O^{17}(n,p)$-N^{17} reaction of the oxygen in the primary coolant. Since the half life of this isotope is 4.14 seconds, a transit time of a minute or so to the DN detecting station allows the N^{17} to decay to a negligible amount.

Failure Location

The multisampling system mentioned in the first section for purposes of failure location can be as extensive as required.[1] In general, the number and location of sampling lines is determined by an analysis of: probable failure zones; the number and location of discreet, replaceable sections of the core; the maximum error in location that can be tolerated; the probable primary system activities that can be tolerated in the event of a failure; and reactor flux distribution. These considerations, among others, also determine the importance attached to a given size of failure. The sampling lines, led through a multiple valve manifold, enable known areas of the core to be surveyed, and the location of the failure is denoted by the highest reading sample line. The line which samples effluent from the region containing a failure will produce the highest reading because of the presence of freshly added DN fission products from the failure location. The other lines will sample the highly diluted activity which returns through the circulating primary system to these other lines. (The relations describing these details have been analyzed and are given in references 2, 3.)

The flux depression method, also mentioned in the first section, is a powerful tool. Eliminating the need for a complex sampling line system, the location procedure can be carried out rather quickly, an advantage arising from the short half-life of the isotopes involved. Only a short wait, of the order of 5 minutes, is necessary for the system to come to equilibrium after a control rod maneuver. The performance of this test demands an approximate knowledge of the flux contours produced by a given rod insertion. A map of flux versus location within the reactor for each rod(s) movement enables one to see the areas of overlap from adjacent rod movements. The extent of the reduction of neutron activity of the primary coolant with a given rod(s) insertion can be correlated with the map for that rod(s) movement. The area of the failure is then located with some precision. By these methods of failure location, costly radiochemical tests are eliminated, and only the fuel in the indicated defective zone need be replaced if replacement is deemed necessary.

The presence of generally distributed surface contamination can be distinguished from a fuel failure by the failure location procedures, which will reveal that no localized source of activity such as a fuel failure is present.

Measurements have been made on the reactor to which this system has been applied to gather information on the nature of the flux plots previously mentioned. These measurements revealed that the flux was depressed an average of 50% in a well defined area about each control rod in moving the rod from the fully withdrawn to the fully inserted position. Little confusion arose from manipulation of adjacent control rods. Although in some cases manipulation of one rod causes measureable flux changes in the region of adjacent rods, the greatest changes are confined to the region of that one rod, so that good resolution is obtained.

Other Reactor Systems

There are no essential differences that exist if one considers this system in terms of air-cooled reactors or boiling-water reactors. Provided that the construction of the reactor enables some form of channeled coolant flow about the fuel units, sampling of air or water can be ac-

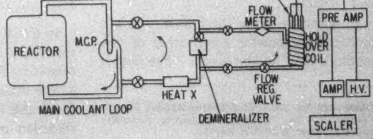


Fig. 2. Outline of DN detection system for fuel-failure detection. This system is also used for failure location with flux depression procedure

Vol.
Pt. 1
1959

complished as previously outlined. A boiling-water reactor, for example, presents no special problems because the reactor core is essentially covered to its full depth with flowing water (a fraction of the upper part of the core contains a steam-water mixture, with water predominant). Any steam included in the withdrawn sample, if sampling procedures are used, would readily condense in the sampling lines; background levels and signal levels would be realtively the same as for nonboiling water systems.

Flux depression techniques, of course, can be applied freely to these other reactor systems, subject to the same criteria.

Description of Equipment

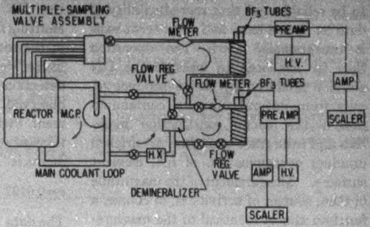
Failure Detection Equipment

The DN detection station which is used is illustrated by Fig. 1. Primarily, it consists of a $3^1/_4$-inch-diameter helical coil of 3/8- by 1/4-inch stainless steel tubing having a total volume of 150 cubic centimeters (cm). Within the coil is a cylindrical polyethylene plug containing three equally spaced holes for the location of neutron counters. Anton Electronics Laboratory type *370-E* proportional counters with 2-mil-diameter anodes are used. The counters are filled with $B^{10}F_3$ to a pressure of 60 cm of mercury and are about 1 inch in diameter and $11^3/_4$ inches in length. A fourth hole located on the axis of the polyethylene plug is used for insertion of a calibration source, nominally 5 millicuries of polonium-beryllium. The source is 1/4 inch in diameter and is attached by a threaded connection to a 1/4-inch-diameter insertion rod. A calibration count rate of the order of 5,000 counts per minute is produced by this source.

In order to shield the counters from ambient neutrons resulting from leakage through the reactor shield, the coil is surrounded by $1^1/_2$ inches of paraffin, 1/8 inch of cadmium, and an additional $6^1/_2$ inches of paraffin.

As shown in Fig. 2, the DN station is connected to the reactor primary coolant system by connection to the coolant purification system. Since the coolant purification loop joins the main coolant system at the inlet and outlet of the primary coolant pumps, a large hydraulic head is available to produce flow through the DN detection station. Also the water in the purification loop is cooled prior to entering the demineralizer and presents the advantage of cooled water for delayed neutron monitoring. The DN detection station is located outside the radiological shielding; therefore, suitable isolation valves are used between the piping outside the shield and the purification loop.

Fig. 3. Outline of DN detection system for fuel failure location by multiple sampling procedure



A high-pressure variable-area-type flow indicater, Brooks Rotameter model no. *FV-1110-VH-BTP*, is used as a reference for regulation of flow to the DN detection station. The flow rate to the station is adjusted to a fixed value which produces a coolant transit time of approximately 2 minutes from the reactor to the DN station. This transit time allows for decay of the short-lived nitrogen-17 neutron activity which is inherently present in water-cooled reactor systems, leaving only bromine-87 and iodine-137 activity present within the coil of the DN detection station.

The counting circuitry is also shown in Fig. 2. It consists of an Atomic Instruments linear preamplifier, model *219;* Atomic Instruments linear amplifier, model *218;* and a Radiation Instruments Development Laboratory scaler, model *49-10.* A Radiation Counter Laboratories count-rate meter, model *15-Mark 15,* and a Foxboro EMF Dynalog recorder are also provided for use when continuous recording is desired. A John Fluke, model *405,* high-voltage power supply is used to provide approximately 2,200 volts to the three counters. Simplex low-noise coaxial cables are used for the high-voltage and input signal cables.

Over-all system gain is set so that the majority of output pulses from the linear amplifier fall in the range of 70–80 volts. This is established with a calibrated scope or a pulse analyzer. The discriminator setting is 20 volts.

Failure Location Equipment (Sampling System)

The equipment used for fuel element failure location by the multisampling technique is auxiliary equipment, installed only at the time of its use. As illustrated by Fig. 3, 19 external sampling lines are connected to 19 sampler receptacles on the reactor pressure vessel. Each sampler receptacle provides a point of connection to a sampler rake installed within the reactor. Each rake provides a representative sample of coolant effluent from a discreet section of the core.

The 19 sampler lines of 3/8- by 1/4-inch stainless steel tubing are led to a multiple valve-manifold assembly which was manufactured by Autoclave Engineers, Inc. This unit permits the selection of coolant effluent from one sampling line at a time for DN monitoring and also regulates the flow rate in the sampling line. Flow from the valve manifold passes through a Graham heat exchanger, model *12X4C-12,* and the cooled water than enters the DN station where the level of DN activity is measured. Outflow from the DN station passes through a Brooks variable area-type flow indicator which is used as a reference for adjustment of flow in the system. From the flow meter, the water then returns to the reactor primary system via a canned-motor centrifugal pump. This pump provides the hydraulic head to obtain a flow rate consistent with a 2-minute transit time of the coolant from the reactor to the DN station. Suitable isolation valves are used to isolate the failure location equipment from the high-pressure reactor primary system.

In operation, the neutron activity of each sampling line is compared to the reactor coolant average activity. Each sampling line is monitored by the DN station for a fixed period of time. The resulting CR observed is compared to the CR produced by the permanently installed primary system DN detection station. Abnormally high neutron activity of a sampler line, compared to the reactor average, indicates the location of the failure within the portion identified with that line.

Results

Failure Detection

Operating experience with the previously described equipment has shown it to be reliable and that reproducibility of data on both a long- and short-term basis is exceptionally good. The operation thus far has been to measure the DN activity produced by fission product recoils from natural uranium contamination in the reactor construction materials and from trace quantities of enriched uranium contamination of fuel element surfaces. Measurement of the magnitude of these sources of activity is, of course, a function almost identical to the measurement of exposed fuel alloy in the core in the instance of a failure.

Microphonic or electronic background with the power plant equipment in operation has been found to be negligible. Background due solely to leakage of environmental neutrons through the detecting station shielding has been approximately 30 counts per minute at full reactor power.

To obtain precision when taking DN measurements on an operating plant, it is necessary that reactor power and coolant transit time between the reactor and the detecting station be maintained constant during each counting interval and that these values be accurately known from one counting interval to another. For the application in question, counting intervals of 15 to 20 minutes were normally used. With meticulous care taken to maintain constant flow and power level, and with standard shipboard gages and flow indicators as references, a fractional standard deviation of less than 3% was obtained in each of two sets of data of 18 and 24 points, taken under two different sets of operating conditions. In another instance, data taken by several different persons unfamiliar with the equipment, under less constant conditions of flow and power level, produced a total of 16 points with a fractional standard deviation of 7%.

Calibration of the measuring system by reference to the decay curve of a nominal 5 millicurie polonium-beryllium calibration source showed less than 1% change in calibration over a period of 4 months.

The absolute counting efficiency, ϵ (see equation 3), of a detecting station was determined in an experiment conducted at the thermal test reactor (TTR) at the Knolls Atomic Power Laboratory.[4] The DN station used in this experiment was of identical construction to that used on the reactor plant, and the results were applied to the measurement made on the pressurized water reactor system. The experiment was designed to provide a continuous recirculation of water containing U-235 in solution through a region of the thermal column of the TTR and thence to the DN detection station. The DN emitting fission product isotopes produced in the reactor thermal column by U-235 fission were therefore carried to the DN detection station where their activity was observed. The data taken in this experiment yielded nine separate determinations of ϵ; the average of these measurements is

$$\epsilon = 0.0127$$

The data had a fractional standard deviation of 2.0%.

If desired, the counting sensitivity of the detecting station described herein could be increased, without significant increase in its over-all size by increasing the diameter of the tubing used in forming the hold-over coil, and by a slight increase in the diameter of the coil. For example, an increase in tubing inner diameter from 1/4 to 3/8 inch would increase the volume of the hold-over coil by a factor of 2.25. An increase in the diameter of the coil to $4^{1}/_{4}$ inches would increase this volume by an additional factor of 1.3 and would also increase the space within the coil for the location of counters and associated moderator by a factor of 1.7. This would permit use of five rather than three neutron counters to monitor the coil. Thus, the over-all sensitivity to coolant delayed neutron activity could be increased by about a factor of 5 with only a 1 inch increase in over-all diameter of the detecting station.

Using the absolute counting efficiency obtained in the TTR experiment, the DN counting data obtained from the pressurized water power reactor were used to calculate the total amount of U-235 in the core which was contributing to the system fission product activity. This amount was compared to the amount determined by radiochemical analyses performed on extracted samples of primary coolant, by which the concentrations of many fission product isotopes were determined. The results, normalized to the value calculated from the DN counts, are shown in the following:

	Relative Amount of U-235
DN value	1.00
Radiochemical value	0.55

The DN value represents one calculation based on the average of 24 measurements with a fractional standard deviation of 3%. The radiochemical value represents the average of 19 individual measurements and calculations based on a total of 11 different isotopes. The fractional standard deviation of the radiochemical values is 60%.

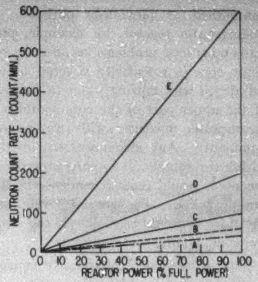


Fig. 4. Typical DN system output from reactor source contributors

A—Ambient background contribution
B—Normal 2 ppm contribution
C—Total background (A+B)
D—Total signal: 1 cm^2 failure
E—Total signal: 5 cm^2 failure

Conditions:
Full power thermal flux 10^4
2-minute delay to detector
2% counting efficiency
400 cm^3 holdover volume
10% U-235 fuel content

To illustrate the relationship and magnitude of the data obtainable from a DN detecting station to other plant parameters, Fig. 4 has been prepared. Fig. 4 is based on the following hypothetical system and counting station parameters:

v = detecting station volume = 400 cm^3
ϵ = detecting station counting efficiency = 2%
V = active primary system volume = 5×10^6 cm^3
ϕ = thermal flux density in core (average) at full power = 10^{14} n/sec-cm^2
f = fraction by weight of U-235 present in fuel alloy = 10%
ρ = fuel alloy density = 5 grams/cm^3
h = fission product recoil range in fuel meat = 10^{-3} cm
S = area of failure (exposed fuel) = 1 cm^2 and 5 cm^2

The foregoing values are substituted in equations 2 and 3 along with the standard constants involved in these equations to calculate the DN activities produced by the counting station. Typical values for counting station background caused by environmental neutrons and to 2 ppm (parts/million) natural uranium contamination in the core mechanical structure are also shown on Fig. 4.

It may be concluded that the DN method of fuel element failure detection is reliable and precise, and that it is a simple technique to apply and operate.

Used in conjunction with sampling systems or flux depression methods, it also represents a very promising approach towards the location of fuel element failures. Since the initial development and application of this system at the Knolls Laboratory, the method has been applied to several other reactor plants.

References

1. FUEL ELEMENT FAILURE DETECTION AND LOCATION, SYSTEM DESIGN DESCRIPTION, J. C. Childs. *Memo SSD 42* (revision 2), Knolls Atomic Power Laboratory, Schenectady, N. Y., 1957 (available from the United States Atomic Energy Commission, Technical Information Service Exchange, P. O. Box 62, Oak Ridge, Tenn.)

2. FUEL ELEMENT FAILURE DETECTION AND LOCATION—ESTIMATE OF SENSITIVITY—DN SYSTEM, R. A. Dewes. *Memo RAD-2*, Knolls Atomic Power Laboratory, 1957 (available from AEC).

3. LOCATION OF MULTIPLE REACTOR FUEL ELEMENT FAILURES BY DELAYED NEUTRON SAMPLING METHODS, R. A. Dewes. *Report 85GL321*, General Electric Company, Schenectady, N. Y., 1958 (available from AEC).

4. DELAYED NEUTRON DETECTION STATION-EFFICIENCY DETERMINATION, R. A. Dewes. *Memo RAD-3*, Knolls Atomic Power Laboratory, 1958 (available from AEC).

The Effect of Flux Distribution on Iron Losses

M. J. SCHINDLER
NONMEMBER AIEE

Synopsis: The classical formula used to compute eddy-current losses in magnetic sheet steel is derived from Maxwell's field equations based on the conditions of constant permeability. This condition of constant permeability is unfounded since in each cycle the permeability of modern materials will undergo changes of at least 1 to 100. Large differences between actual losses and the losses computed according to the classical formula have been noted. These differences could be attributed to the assumption of constant permeability. This paper attempts to determine approximate solutions for Maxwell's field equations under the condition of variable permeability that will lead to a modified formula for determining the eddy-current losses. Since rough approximations were used, final conclusions cannot be made. However, the results indicate that the error in the classical formula (the assumption of constant permeability) accounts for only a small part of the discrepancy between calculated and actual eddy-current losses. Experiments were conducted that confirmed the concept of changing flux distribution. These experiments also indicated that other mathematical approaches could be used for more accurate approximations.

THE losses in magnetic materials at low frequencies are generally divided into two parts: the hysteresis loss and the eddy-current loss. The hysteresis loss is attributed to the energy consumption of the flux reversal and is determined by plotting the hysteresis loop at a very low frequency.

The eddy-current loss is due to resistive heating by eddy currents, and can be computed if the flux distribution and the material properties are known throughout the magnetization cycle. In the past, the flux distribution has been found by solving Maxwell's field equation on simple geometric configurations based on the assumption of constant permeability. The solution of these equations is known as the classical formula.

This approach has resulted in very considerable differences between the calculated and actual figures; Stewart in one instance found the eddy-current loss to be three times the calculated value.[2] This difference is called the eddy-current anomaly. However, it is based on a formula known to be incorrect. Therefore, it seems reasonable first to estimate where the classical formula is in error and in what direction. It appears that the error may be caused by basing the computations on constant permeability.

This paper attempts to find approximate solutions for Maxwell's field equation under the condition of variable permeability that will lead to a modified equation for the eddy-current losses.

Flux Distribution

The classical formula has been derived under the condition of equal flux distribution across the thickness of the sheet. Whether this assumption holds true can easily be estimated by computing the so-called penetration depth. Since, as far as eddy currents are concerned, the most critical moments during the magnetization cycle are the ones corresponding to the steep parts of the hysteresis loop, it is necessary to use the maximum value of the differential permeability in the penetration equation

$$\delta_{[cm]}=50.33\sqrt{\frac{\rho}{\mu f}}$$

For hot-rolled silicon steel this approach yields figures between 1,000 and 5,000, leading to penetration depths of 0.5 to 1 mm (millimeter).

Because the thickness of normal transformer sheets lies between 0.3 and 0.5 mm, equal flux distribution can be assumed and the classical formula should give fairly accurate values. However, modern textured materials have loop slopes of up to 10^6 or more. A typical value (Armco Trancor *3X*; see Fig. 1) is 3×10^6 which gives a penetration depth of approximately 0.04 mm. According to the classical theory, the magnetic field decreases exponentially with

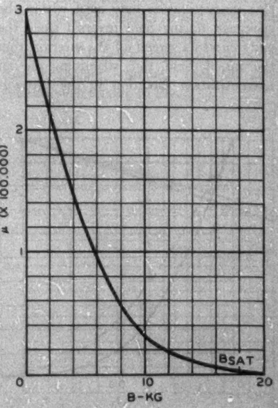


Fig. 1. Differential permeability versus induction for Armco Trancor 3X, hysteresis eliminated

---

Paper 59-1204, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted April 9, 1959; made available for printing August 26, 1959.

M. J. SCHINDLER is with the Radio Corporation of America, Harrison, N. J.

The author wishes to express his gratitude to the Wright Air Development Center for the opportunity of working on this project, as well as for permission to use the material for this paper, which is based upon work done while the author was employed at the Metallurgy Branch, Aeronautical Research Laboratory. The material, containing the mathematical deductions in detail and a description of the low-frequency generator used for the experiments, was reported in 1958.[1]

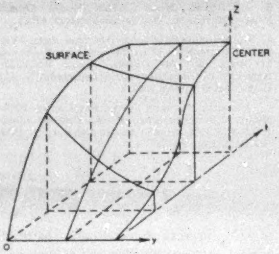


Fig. 2. Flux variations for sinusoidal flux change and exponential distribution

$$H_{(Y)}=H_0E^{-y/\delta}$$

In the center of a sheet 0.16 mm from the surface, the field would therefore be only 0.02 of the surface field. This result is definitely far enough from homogeneous to render the classical theory invalid.

To describe the actual process more accurately, it is necessary to compute the eddy-current losses under the condition of nonhomogeneous flux, exponentially decreasing toward the center plane. But this is not enough. Only during the time of magnetization reversal, when the induced eddy-current voltage is highest, can this high-screening effect be expected. As the material approaches saturation, the flux becomes more and more equally distributed over the whole cross section as the differential permeability decreases. This flux distribution causes the penetration depth to decrease. Furthermore, as Rusbüldt[3] points out, positive and negative magnetization may exist within the specimen during field reversal since the eddy currents screen off the interior from the outside field, thus sustaining the old magnetization direction in the interior. In the extreme, this situation would lead to a model in which two huge domain walls move from the surface toward the center, reversing the induction from plus to minus saturation, or vice versa, as they pass. The true state lies somewhere between these extremes, but where cannot be predicted without solving the field equations.

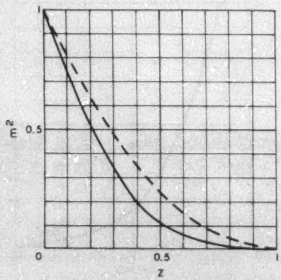


Fig. 3. Solid curve: normalized curve of Fig. 1. Dotted curve: approximation by m=1−Z

Exponential Flux Distribution

GENERAL CONCEPT

Since the distribution of flux or induction across the thickness is not known, another assumption must be made. The new theory should come closer to actuality than does the classical theory if it is assumed that the permeability will be uniform across the whole sheet at any given moment, but that the permeability will be a function of the average induction at this time. In other words, the flux decrease toward the center plane will be highest at low inductions and will approximate zero with saturation, where the differential permeability becomes very small. Of course, in a model of this kind there will be one direction of flux only, at any time, over the whole cross section, and the simultaneous existence of both directions at the time of flux reversal cannot be accounted for.

In the case of sinusoidal induction, this approach would lead to a distribution over time and cross section, as shown in Fig. 2, causing formation of harmonics on the surface as well as in the center region and a higher average induction on the surface than in the center region.

To facilitate computation it is necessary to find a suitable approximation for $m=\sqrt{\mu}$ as a function of the induction B. Fig. 1 shows u as a function of B for Armco Trancor *3X*, showing the maximum of 300,000 at $B=0$, and the approach toward zero at saturation induction B_{sat}. To make this function independent of the absolute values of B and μ, it is convenient to introduce a "normalized" induction $z=B/B_{sat}$ and replace the relative permeability by m^2. This step immediately points toward a very simple approximation of the function, namely

$$m=1-Z$$

which, as Fig. 3 shows, comes fairly close to the measured values.

COMPUTATION OF EDDY-CURRENT LOSSES

To find the energy consumed by the eddy currents in unit volume, it is necessary to compute the differential energy in an infinitely thin layer and integrate over

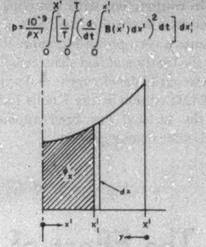

$p=\frac{10^{-9}}{\rho X'}\int_0^{X'}\left[\frac{1}{T}\int_0^T\left(\frac{d}{dt}\int_0^{x_1'}B(x')dx'\right)^2dt\right]dx_1'$
ϕ_x
dx
x'
x_1'
x'
y

Fig. 4. Flux contributions to eddy currents in plane x_1'

the sheet thickness, dividing the result by the thickness. For reasons of symmetry, however, only one half of the sheet is considered, integrating from the center layer toward the surface. This procedure can be simplified by using x' instead of y for the local variable, counting x' from the center plane rather than from the surface; see Fig. 4.

The power per unit volume will be

$$p=\frac{10^{-9}}{\rho X'}\int_0^{x'}\left[\frac{1}{7}\times\int_0^T\left(\frac{d}{dt}\int_0^{x_1'}B(x')dx'\right)^2dt\right]dx_1'$$

or, using $B=B_{max}\sin\omega t$ and D for the total thickness $2X'$:

$$p_c=\frac{10^{-9}}{6\rho}\pi^2f^2B^2_{max}D^2$$

which is the classical formula for the eddy-current losses.

For the present purpose, the normalized induction z is substituted and, to simplify computation, a modified local variable $x=x'/\Delta$. This substitution

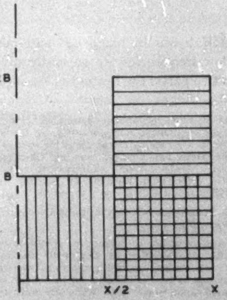


Fig. 5. Vertically hatched area—uniform flux distribution. Horizontally hatched area—flux concentration on surface

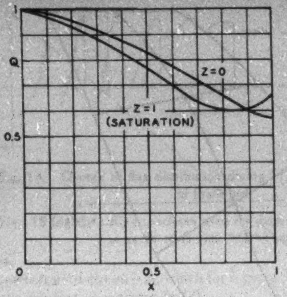


Fig. 6. Loss ratio for improved linear approximation

means that the thickness is measured in penetration depths rather than in centimeters thereby making it independent of the geometrical configuration. The final equation for the losses then reads

$$p=\frac{10^{-9}}{\rho}\Delta^2 B_s^2 \frac{1}{X}\times \int_0^{x}\left[\frac{1}{T}\int_0^T\left(\frac{d}{dt}\int_0^X z\,dx\right)^2 dt\right]dx$$

This equation shows that, if the time function z is a sequence of harmonics, the time integral will be the sum of the integrals of its members. Therefore, an increase must be expected in the loss figures if harmonics are generated; see Fig. 2.

It is now possible to calculate roughly what influence the second factor, namely, higher average induction in the surface layer, will have on the total loss figure. This calculation can be made easily by comparing the losses for the two different cases of flux distribution given in Fig. 5, where the vertically hatched area is a homogeneous flux which leads to the classical formula, while the horizontally hatched area denotes a distribution that would be the extreme of the situation in Fig. 2. Both areas are equal, and since a pure sinus is used as time function in both cases, there will be no difference in the magnetic behavior of the two sheets. As long as the induction $2B$ is not higher than saturation, it is not known which situation really exists.

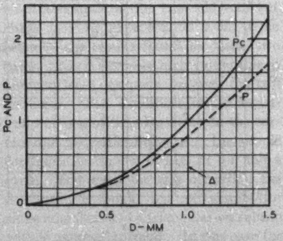


Fig. 7. Loss figure p (nonuniform flux) compared with classical result p_c

To compute the loss for the second case (concentration of the flux in the surface region), the lower integration limits are changed from 0 to $X/2$, and the induction B is replaced by $2B$. The former introduces a factor of 1/8, the latter of 4, giving a final result of $p=1/2\ p_c$. This result (decrease of the losses to half the original value) may, at first sight, look surprising. However, it is quite evident that the flux in the center region contributes most to the eddy currents since a flux path using two symmetrical planes at the distance d from the center will induce a voltage only in layers with greater distances than d, while the space between the flux-carrying layers will be free from eddy currents.

These preliminary considerations may be summed up by stating that there will be a loss increase due to the harmonics and a decrease due to the unequal flux distribution without evidence as to which will be greater.

LINEAR APPROXIMATION

The loss computation for an exponential flux distribution turns out to be impractical. A fairly simple approximation must be used, and for better understanding of the procedure a straight line is chosen which corresponds to the first two terms of the sequence for the exponential function

$$\epsilon^{-mx}=1-mx+\frac{1}{2}m^2x^2\ldots$$

Following the steps outlined above reveals that

$$p=\frac{10^{-9}}{6\rho}\pi^2 f^2 Z_0^2(2\Delta X)^2 Q=p_c Q$$

Consequently, the result can be written as the classical formula multiplied by a factor Q. Since the classical formula will give correct results if the sheet is thin enough, this conclusion was expected. The factor Q is given by

$$Q=1-\frac{1}{4}X-0.23X^2+0.05X^3+0.03X^4+ Z_0^2(-0.45X^2+0.15X^3+ 0.43X^4)+0.025X^4Z^4$$

which is plotted in Fig. 6.

This result can be approximated roughly by $Q=1-0.3X$ resulting in a modified formula for the eddy-current loss

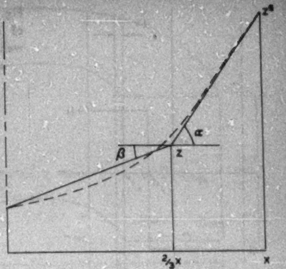


Fig. 8. Composed linear approximation of exponential flux decrease

$$p=\frac{10^{-9}}{6\rho}\pi^2 f^2 B^2 D^2\left(1-0.3\frac{D}{2\Delta}\right)$$

This equation was compared with the classical formula in Fig. 7 for a Δ of 1 mm and, although it definitely gives no account for the eddy-current anomaly, the modified equation points out that with varying sheet thickness the eddy-current loss will not change as much as the classical formula predicts.

This has been observed before but it was usually attributed to surface layers of allegedly poorer magnetic qualities. This explanation may be true to some extent but since the penetration depth usually is of the same order of magnitude as the sheet thickness (no one uses thinner material than is strictly necessary), the conditions of Fig. 7 will usually prevail. It must not be forgotten that p is the loss factor per unit volume which should be expected to decrease, if only part of the cross section is being magnetized, as is the case in thicker laminae. (See the considerations referring to Fig. 5.)

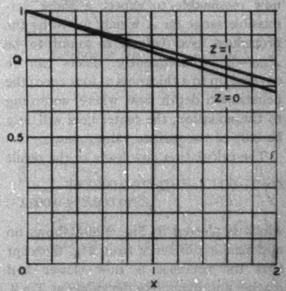


Fig. 9. Loss ratio for composed linear approximation

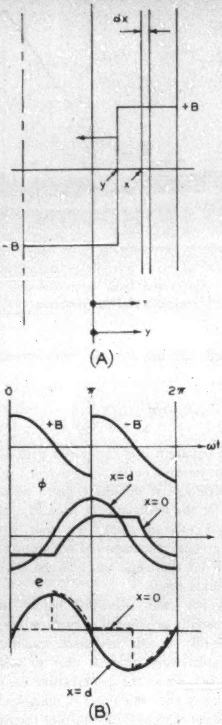


Fig. 10 (left). Movement of a single huge domain wall, and resulting flux and eddy-current voltage

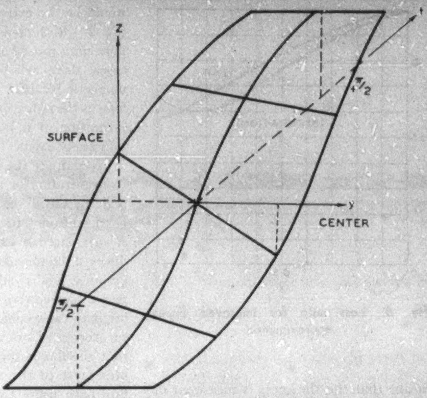


Fig. 11 (right). Flux variation for sinusoidal flux change and "swinging" distribution

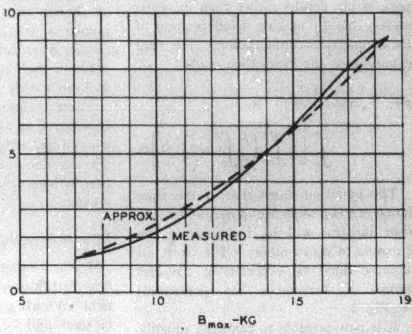


Fig. 12 (right). Hysteresis loss versus maximum induction for cold-rolled steel

COMPOSITE LINEAR APPROXIMATION

Admittedly the linear approximation used will not describe the exponential induction decrease very well. A better approach may be to divide the model into two parts using a different linear approximation in each one, as shown in Fig. 8.

Due to the better approximation, it is now reasonable to expect a wider permissible range for X which, following the procedure used before, is found to be $X_{max}=2$. Consequently, sheets can be covered up to a thickness of four times the penetration depth now, where, according to the equation, the center field will only be 0.02 of the surface value.

The calculation now leads to the result

$$Q=1-0.11X-0.03X^2+0.002X^3+Z_0^2(0.006X^2+0.002X^3)$$

which is plotted in Fig. 9 and shows no appreciable difference from Fig. 6 except that the decrease is now slower and reaches approximately the values at $X=2$, as was found for $X=1$ in Fig. 6. This result points out that the approximation is not very critical as long as the general distribution between zero and maximum is covered.

The conclusion so far is that an exponential decrease of the induction from the surface of a sheet toward the center will not lead to an increase in eddy-current losses and that a different approach is necessary.

Movement of a Huge Domain Wall

As mentioned in regard to flux distribution, it is reasonable to expect magnetization of the positive and negative signs simultaneously during field reversal which in the extreme would lead to a model in which two huge domain walls move from the surface toward the center. The nomenclature of Fig. 10 has been chosen to simplify computation and it therefore deviates somewhat from that used in the previous sections.

The thickness of the model would be $4d$ but only one half of the cross section is shown in the figure. The surface is at the right side where the domain wall movement starts with $y=d\cos\omega t$, giving a total flux as depicted in Fig. 10(B), if the local variable x equals d. In any other location the flux will change only during part of the cycle, as shown for $x=0$, remaining constant until the domain wall enters the region for which it is to be computed. This is the case at the time

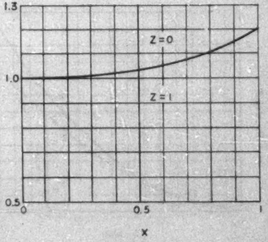


Fig. 13. Dynamic hysteresis loss versus sheet thickness

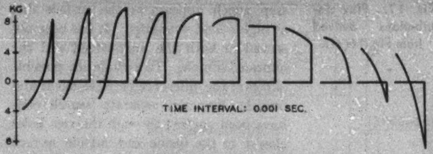


Fig. 14. Change of flux distribution during 1/2 cycle in sheet model of Brailsford

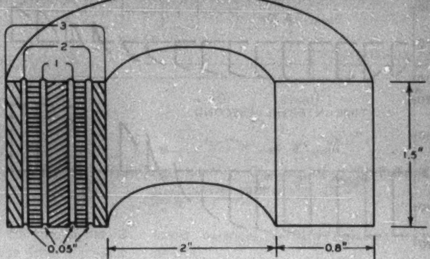


Fig. 15 (right). Ring specimen used for determining flux distributions as depicted in Figs. 16(A) and (B)

t_1 when $y = d \cos \omega t_1 = x_1$, which for t_1 gives the solution $\omega t_1 = \arccos (x_1/d)$.

For any layer, therefore, it is necessary to compute the time integral between this value and $t = 1/4f$ only, because no voltage will be induced at any other time; see Fig. 10(B).

Following the same steps used in the previous sections, it is found that

$$p = \frac{1}{2d}\int_{-d}^{+d} dP = \frac{10^{-9}}{\rho}\, 4\pi f^2 B^2 d \times$$

$$\int_{-d}^{+d}\left[\pi - \arccos\frac{x}{d} - \frac{x}{d}\times\right.$$

$$\left.\sqrt{1-\left(\frac{x}{d}\right)^2}\,\right]dx = \frac{10^{-9}}{\rho}\, 4\pi^2 f^2 B^2 d^2$$

and, as the classical formula for a thickness of $4d$ gives $P_c = (10^{-9}/\rho)\pi^2 f^2 B^2 (4d)^2$, the final result reads

$$Q = \frac{p}{p_c} = 1.5$$

Although this result would indicate higher losses as compared to the classical formula, it must not be forgotten that no commercial sheet material provides the conditions under which a huge single wall would form. Rather, it must be expected that the true figure will lie somewhere between this result and the one found previously.

Swinging Flux Distribution

Knowing that the true flux distribution would have to lie between the two previous assumptions, the true eddy-current losses may be expected, as well, to lie between the results found for these two cases. Still, it would be highly desirable to know at least one point between the two extremes. The condition would essentially be one of time lag between the surface and the center fluxes, e.g., the simple configuration shown in Fig. 11 for half a cycle. But even for this condition the computation would not be simple for sinusoidal flux, and a triangular waveform will be assumed instead. In this case the induced voltages are constant, which simplifies the problem considerably. On the other hand, the deviation from the sinus is not very great and any change in loss values, due to different flux distributions, should be about the same in both cases.

It will be assumed that the slope of the distribution line changes sinusoidally, and the steps used in the earlier calculations will be followed. It is found that

$$p = \frac{16}{3}\, 1.37 Z_0^2 f^2 X^2$$

which must be compared to the result for $z = Z = Z_0(4ft-1)$. This is simply given by the first member of the solution for p, being

$$p_c = \frac{16}{3} Z_0^2 f^2 X^2$$

thus leading to the final result $Q = 1.37$.

Dynamic Hysteresis Loss

One fact often overlooked is that the hysteresis loss, too, may change under a-c conditions. If, for instance, exponential flux distribution is assumed, the center parts of the sheet will never reach the induction which would be expected from the total flux, while the surface layers will reach higher values. To calculate this error, a mathematical expression for the hysteresis loss as a function of induction is required. Fig. 12 shows this function for a cold-rolled material. In the practical range of between 7 and 17 kilogausses it follows fairly accurately a square law.

Introducing the flux distribution into the formula for the hysteresis loss

$$p_h^* = \frac{1}{X}\int_0^X L_h Z_{\max}^2 dx$$

reveals that

$$p_h^* = \frac{1}{X} L_h Z_0^2\left[X + (1-Z)^2\left(1+\frac{1}{2}X\right)^2 \times\right.$$

$$\left.\left(\frac{1}{3}X^3 - \frac{1}{2}X^3 + \frac{1}{4}X^3\right)\right]$$

or, because the static or d-c hysteresis loss $p_h = L_h Z_{\max}^2$,

$$p_h^* = p_h\left[1 + \frac{1}{12}(1-Z)^2\left(1+\frac{1}{2}X\right)^2 X^2\right]$$

$$= p_h Q_h$$

Q_h is plotted in Fig. 13. The maximum possible increase of the hysteresis loss would be less than 20%, and this would be at very low inductions and for a sheet in which the center is practically completely free from magnetic flux.

Experimental Results

To simulate conditions in a magnetic sheet, Brailsford[4] built a model using a stack of transformer steel strips, on which he put loaded windings to act as eddy-current paths. He found that, although the total flux in the stack was sinusoidal, the flux in the center appeared peaked, while on the surface the curve was flattened. This result is in accordance with Fig. 2. Furthermore, a time delay

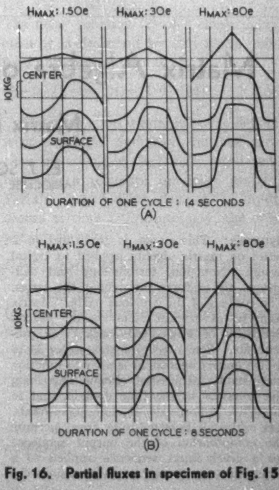


Fig. 16. Partial fluxes in specimen of Fig. 15

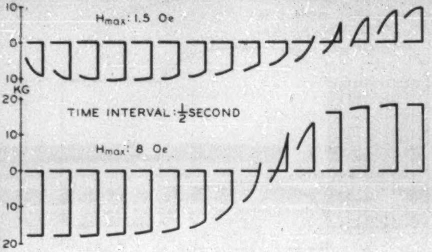


Fig. 17. Flux distributions derived from Fig. 16(A)

in the center part can be observed. From his curves the flux distribution at different moments have been derived in Fig. 14. They show clearly the generally logarithmic lag toward the center of the sheet during flux changes, and the simultaneous existence of positive and negative inductions. The sequence shown in Fig. 14 more or less resembles a damped traveling wave and could probably be treated mathematically as such.

While at power frequency only a model can be used to gather information about the flux distribution, this detour may be avoided by using very low frequencies. Fig. 15 shows part of a ring through which holes of 0.05-inch diameter have been drilled in such a way as to make the differently hatched areas equal. Three search coils of 30 turns each have been wound as indicated by the braces 1, 2, and 3. Coil 1 will then give the flux in the central part. The differential voltage between 2 and 1 gives the flux in the intermediate section and, for the surface layer, 3 is connected in counterphase with 2. This sample has been magnetized by a very-low-frequency generator capable of delivering 500 *W* peak triangular current at less than 0.05 cycle per second. The output voltages of the search coils have been integrated electronically, with the use of a Reeves d-c amplifier with capacitive feedback, and plotted on a Librascope.

Fig. 16 displays the flux in the three sections of the sample (center, intermediate, surface) at three different field values. The duration of one cycle is approximately 14 seconds in Fig. 16(A) and approximately 8 seconds on Fig. 16(B). From these, Fig. 17 has been derived showing the flux distribution at time intervals of one second. They are in agreement with those of Fig. 14, if the triangular field is considered against the sinusoidal flux used by Brailsford.

A few words should be added about the step which appears in the surface flux at higher field strengths. It has been observed in both the unannealed and the annealed sample. To exclude any possible effects of the difference between inside and outside diameter, separate search coils have been applied through the two holes closest to the inside and outside surface and the flux plotted for both. Although a time lag can be observed because of the higher inside field strength, the step is present in both cases. It is caused by the fact that the surface layer has already reached a flat part of the hysteresis loop when the center enters the steep portion, and therefore takes over practically all the change in the total flux. Only when the center, too, has reached the flat part of the loop will the surface flux increase simultaneously with the center flux, as may be seen from the charts. This effect will be emphasized by a rectangular loop. A noticeable increase has been observed after annealing of the sample.

References

1. The Effect of Flux Distribution on Iron Losses, **M. J. Schindler.** *WADC technical note 58-174*, "ASTIA document no. AD203785," Armed Services Technical Information Agency, Washington, D. C., 1958.

2. Losses in Electrical Sheet Steel, **K. Stewart.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 97, pt. II, 1950, pp. 121–25.

3. Die Wirbelstromanomalie bei kornorientierten Elektroblechen, **V. Rusbüldt.** *Elektro-Technische Zeitschrift*, Wuppertal-Elberfeld, Germany, vol. 78, 1957, pp. 337–42.

4. Investigation of the Eddy-Current Anomaly in Electrical Sheet Steel, **F. Brailsford.** *Journal*, Institution of Electrical Engineers, vol. 95, pt. II, 1948, pp. 38–98.

Vol. 7
Pt. 1
1959

Matrix Algebra of Sequential Logic

Matrix Logic III

E. J. SCHUBERT
ASSOCIATE MEMBER AIEE

Synopsis: This paper is the third in a series on matrix algebra applied to logical design. Notation, terminology and definitions are uniform throughout the series. The content of the previous parts, "Matrix Analysis of Logical Nets"[1] and "Matrix Synthesis of Highspeed Logic,"[2] is prerequisite. The object of this paper is to derive matrix operations similar to function matrices for sequential circuits describing the behavior in the time domain, and to introduce appropriate algebraic operations. The approach, believed to be novel, permits the computation of the output sequence for a given input sequence applied to a sequential circuit and in its inversion, the computation of a function matrix required to produce a certain output sequence from a given input sequence.

DIGITAL DATA processing implies the extensive use of sequential logic, which is a combination of static logic and memory functions. An output from static logic depends solely on the combination of inputs irrespective of time. Memory functions in the abstract sense imply the preservation of a state or a signal for a certain length of time without any logical operation on the signal. Feedback loops may be provided to control the inputs or the outputs of memory elements, depending on previous states.

Static or combinatorial logic is frequently represented in the space domain as a Venn diagram or in its abstracted form called a Veitch-Karnaugh chart. This approach becomes impracticable with increasing complexity. Because any logical proposition may be expressed as a system of linear forms in canonical minterms, matrix algebra has been advanced in two previous papers.[1,2] This paper is

---

Paper 59-204, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted January 23, 1958; made available for printing July 3, 1959.

E. J. Schubert, formerly with the Burroughs Corporation, Paoli, Pa., is now with Monitor Systems Inc., Fort Washington, Pa., subsidiary of Epsco, Inc.

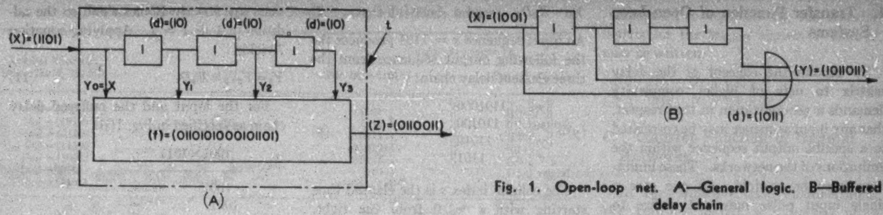


Fig. 1. Open-loop net. A—General logic. B—Buffered delay chain

the third part in the same series and deals with sequential logic. It is written in direct continuation of the previous papers in order to avoid repetition.

Logical multiplication and solution of simultaneous equations by a degenerated type of matrix inversion are feasible. The simplicity of the matrix algebra in sequential logic appears to be advantageous over other presently known concepts.[3–5]

1. Information Content of a Sequential Word

Any input sequence (x_q) may be expressed by a matrix wherein $_q$ refers to the elapsed time increments. The length of a sequence has an upper bound p depending on the over-all system but within this limit it may be defined by the number of bits from the least to the most significant one. The implication of this definition is that a specified sequence of zeros as such cannot be recognized in a single channel unless at least a single ONE has been transmitted as a start signal, and sampling at a given rate is provided.

This fact influences directly the information content of a sequential word of length p. The number combinations for a word with p digits radix b are b^p, e.g., for $b = 2, 2^p$. For sequential words, only those combinations are valid which are distinguishable with respect to the least significant ONE occurring first, i.e. a sequential word of p bits yields 2^{p-1} combinations only.

Because of this reduced information content of an input sequence length p there are not more than 2^{p-1} distinguishable output sequences. This upper limit cannot be exceeded independently of the logical design as long as the coding is binary.

2. Time Differential of a Function

The change of a logical function in time becomes of such significance in this context that a more specific symbol is desirable. Referring to the signal Z in particular the change from its value Z_1 to Z_2 during the transition from time increments T_1 to T_2 can be formulated even for an infinitely short transition from T_1 to T_2. The change of Z may be expressed by a difference Z_2-Z_1, considering that there are three alternatives.

No change:

$$Z_2=Z_1 \qquad \dot{Z}_1=\dot{Z}_1{}'=0 \tag{1A}$$

Switch on:

$$Z_1=0;\ Z_2=1 \qquad \dot{Z}_1=1 \tag{1B}$$

Switch off:

$$Z_1=1;\ Z_2=0 \qquad \dot{Z}_1{}'=1 \tag{1C}$$

To describe the transition of a binary signal requires a ternary notation accordingly. Such a difference Z_2-Z_1 with respect to an infinitely short interval resembles a differential quotient of analytic functions and may be denoted by a similar symbol $\dot{Z}$.[6] If a time index for $\dot{Z}$ is required, the suffix of the first time increment shall be used.

To employ binary coded ternary notation the following convention is established:

States		Binary Coding		Ternary Notation
Z_1	Z_2	$\dot{Z}_1{}'$	$\dot{Z}_1$	
0	0	0	0	0
0	1	0	1	1
1	0	1	0	2
1	1	0	0	0

(2)

States at time T_1, T_2 are related to the differentials and any general logical condition as follows:

$$Z_2=Z_1{}' \quad \dot{Z}_1=Z_1{}' \quad A$$

$$Z_2{}'=Z_1 \quad \dot{Z}_1{}'=Z_1 \quad B \tag{3A}$$

By cancelling identical terms the differentials may be obtained directly:

$$\dot{Z}_1=A$$

$$\dot{Z}_1{}'=B \tag{3B}$$

If $\dot{Z} + \dot{z}'$ is referred to, particularly in larger matrices, ternary notation may be used.

3. Delay Matrix

Sequential logic may be treated somewhat similarly to combinatorial logic as linear forms in the time domain. Corresponding to the minterms in combinatorial logic are time increments. The latter are denoted by an index e for the elapsed time with respect to a zero reference time. The behavior of a delay element is described by the output signal Z_s during the time increment s, depending on the applied input sequence X_q:

$$Z_s=X_qD_e \quad (z)=(x)(d) \quad s=e+q \tag{4}$$

The delay circuit is completely defined by the delay matrix (d) as the output signal following a unit pulse at time zero. Because of the digital nature the output levels are discrete and restricted at present, to binary signals.

To refer to a specific example, in Fig. 1 three delay elements are provided in a chain similar to a shift register with the individual outputs (y_1), (y_2), (y_3). The delay function (d) of a single delay element is described by two terms only for a unit pulse at time 0:

$$D_0=0;\ D_1=1 \quad (y_1)=(1\ 0)=(d)$$

The convention is established to denote with the least significant bit the zero time increment in the binary word representing a sequence or delay matrix. For the second delay element the input occurs at time 1. Accordingly, the output signals are

$$D_0=0;\ D_1=0;\ D_2=1 \quad (y_2)=(1\ 0\ 0)$$

It will be noted that the sequence (y_2) must be obtained according to equation 4 as the product of an input sequence (y_1) in a delay element (d).

$$(y_2)=(y_1)(d)=(d)^2;\ (y_2)=(1\ 0)^2=(1\ 0\ 0)$$

Cascading the signal (y_2) into a third delay element (d) eventually yields the output y_3:

$$(y_3)=(y_2)\ (d)=(d)^3 \quad (y_3)=(1\ 0\ 0\ 0)$$

4. Transfer Function of Open-Loop Systems

To expand the concept of the delay matrix to nets of higher complexity demands a generalization in that respect, that any input sequence may be correlated to a specific output sequence within the limitations of the networks. These limitations may be twofold. In one case a single input pulse may not suffice to generate any output. In the other extreme the length of the input sequence might considerably exceed the number of delay elements. A sequence may be regarded as a vector in binary notation, and accordingly the affine transformation of the input sequence into a specific output sequence implies a tensor of somewhat degenerated form. This transfer function is closely related to a composite truth table with time increments as added parameters.

Transitions of states and outputs may be combined into a single matrix describing the transfer function of a net. Similarly to the continuous relation between input and output in servosystems, the transfer function of a digital system defines the output sequences as a function of input sequences. To derive the matrix of a transfer function the following approach is contemplated.

The concept of the transfer function may be derived from the elementary case shown in Fig. 1(A). The network comprises several delay elements d connected to a combinatorial logic with the function f but without any feedback loop. This latter feature is significant and emphasized by the designation of an "open-loop net."

In this example the static logic is combinatorial for four inputs and is completely described by the function matrix f.

$$Z_s = N_{si}F_i \quad (z)=(n)(f) \tag{5}$$

N_{si} are the combinations of the delayed inputs y, depending on the input sequence x and the elapsed time. The set of intermediate outputs is described by the delay matrix d representing any shift register or delay chain by the unit matrix.

$$(d) = \left\|\begin{matrix} 1 & 0 & 0 & 0 \\ & 1 & 0 & 0 \\ & & 1 & 0 \\ & & & 1 \end{matrix}\right\| = I \tag{6}$$

The first pulse X_0, passing through the delay chain, causes the outputs as in equation 4. Any following input X_q will be shifted in time and will accordingly generate outputs y with increasing delay e, depending on the length of the chain.

$$Y_{es} = X_q D_e \quad s=q+e \quad (y)=(x)(d) \tag{7}$$

An input sequence $x = 1101$ produces the the following output sequence from the three element delay chain:

$$(y) = \left\|\begin{matrix} y_3 \\ y_2 \\ y_1 \\ x \end{matrix}\right\| = \left\|\begin{matrix} 1101000 \\ 110100 \\ 11010 \\ 1101 \end{matrix}\right\| \tag{8}$$

The column index s is the elapsed time starting with $s = 0$ from the right. The row index e designates the output line of the delay chain. Each column represents a specific input combination N_{si} during the increment s, wherein i is the binary word formed by the column with the least significant bit at the bottom. According to previously established conventions, the dummy index i is expressed in octal notation representing as matrix the sequence of minterms at the input of the static logic. $(y)^*$ denotes the transpose of (y).

$$(i)=(10, 14, 6, 13, 5, 2, 1)=(y)^* \tag{9}$$

An output Z_s is restricted to those combinations N_{si} of the intermediate variables Y_s for which the logic is operative; therefore $F_i = 1$. For any possible input sequence (x) the appropriate sequences (i) are essential in order to obtain (z). Because of the restrictions for (x) as stated in section 1, 2^{p-1} combinations have to be evaluated only. According to equations 7 and 9 the inputs into the logic (f) are described by the minterms N_{si}, wherein (i) may be computed directly from equation 10 with the use of a suitable subroutine depending on the digital machine available.

$$(i)=(xd)^* \tag{10}$$

The computer routine requires only that the function matrix (f) is stored internally or accessible at the input for any (i) in order to print the output sequences (z). Input sequences may be generated with another subroutine for the binary expression:

$$x=2n+1 \quad 0 \leqq n \leqq 2^{p-1} \tag{10A}$$

A special case of considerable significance is a chain of delay elements feeding a buffer. The delay matrix d is by definition the output sequence for a unit pulse at zero time. Because no gates are contained in the circuit and no other signal is fed into the buffer than the taps of the delay chain the delay matrix is sufficient to define the circuit. The shift in time for each subsequent input X_e corresponds to the shift operation in a multiplication. No carry as in binary arithmetic is applicable because all operations are strictly Boolean and so the addition of $1 + 1 = 1$. Applying equation 7 yields

$$Y_s = Y_{q+e} = X_q D_e \tag{11}$$

For the input and the buffered delay chain as specified in Fig. 1(B)

$$\begin{array}{r} 1001\times1011 \\ \hline 1001 \\ 1001 \\ 1001 \\ \hline \end{array}$$

$$(y)=1011011$$

5. Closed-Loop Systems

The previous section points out that the input sequence passes through a delay chain which virtually acts as a buffer to store the input sequence in parallel fashion at the inputs into combinatorial logic. In this sense the network behavior was evaluated without internal storage elements.

Closed loop systems imply feedback. A feedback loop is a connection from an output of a logical net to an input of the same net. In this context the loop comprises delay or storage elements but no logical elements. The outputs of these internal storage elements define the state of the net similarly to the Moore-Mealy concept.[3,4] Transition from one state to another is possible in two different modes either requiring a separate reset signal as for a flip-flop element, or without reset as in dynamic delay elements similar to those used in the National Bureau of Standards Eastern Automatic Computer and described by a delay matrix. To discriminate between the two different types of storage elements and the mode of transition, the first one will be called "static storage" and the second one "dynamic storage" or, briefly, "delay."

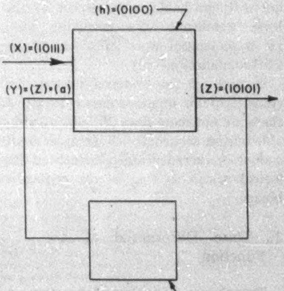


Fig. 2. Dynamic feedback

Table I. Operative Logic with Single Dynamic Feedback Loop

Logical Function (h) = (H₀ H₁ H₂ H₃)	(x) = (1)	Output Sequence (z) For (x) = (:01:)	Output Sequence (z) For (x) = (:1:)
$(h_2) = (0100)$	1	(:01:)	(:01:)
$(h_5) = (1010)$	111110	(:10:)	(:0:) trivial
$(h_6) = (0110)$	(:1:)	(:0011:)	(:10:)
$(h_8) = (1110)$	(:1:)	0101011	(:01:)
$(h_{11}) = (1001)$	(:10:)	(:0110:)	(:0:)
$(h_{12}) = (0101)$	1	(:01:)	(:1:) trivial
$(h_{13}) = (1101)$	(:01:)	(:01:)	(:1:)

Some elementary requirements for the design of such nets may be derived from a basic circuit. In the primitive case of Fig. 2 a sequential input (*x*) and the output (*y*) of a dynamic delay element is connected into a static logic with the function (*f*). Its output (*z*) is fed back into the delay element (*d*). To compute the output signal (*z*) the expansion theorem[2] may be used. A logic with two inputs (*x*), (*y*) is restricted to one of sixteen function matrices (*h*); see equation 2(C) of reference 2. The primitive circuit has two states only defined by the output signal Y_e of the delay element. Accordingly, the output signal (*z*) is correlated to the same states due to the feedback loop.

$$Z_e = f(H, X_e, Y_e) \tag{12A}$$

$$Y_e = Z_{e-1} \text{ for } (d) = (10) \tag{12B}$$

$$Z = H_0X'Y' + H_1XY' + H_2X'Y + H_3XY \tag{12C}$$

$$Z = Y'(H_0X' + H_1X) + Y(H_2X' + H_3X) \tag{12D}$$

Substituting for *Y* equation 12(B) and referencing in time yields

$$Z_e = Z'_{e-1}(H_0X_e' + H_1X_e) + Z_{e-1}(H_2X_e' + H_3X_e) \tag{12E}$$

The equation may be split into two wherein the first describes the condition for switching the signal (*z*) ON, and the second determines the condition to switch (*z*) OFF. For the latter relation some transformation is required.

$$Z_e = Z_{e-1}'(H_0X_e' + H_1X_e) \tag{12F}$$

$$\begin{aligned} Z_e' &= Z_{e-1}(H_2X_e' + H_3X_e)' \\ &= Z_{e-1}[(H_2' + X_e)(H_3' + X_e')] \\ &= Z_{e-1}(\cancel{H_2'H_3'} + H_2'X_e' + H_3'X_e + \cancel{X_eX_e'}) \end{aligned}$$

$$Z_e' = Z_{e-1}(H_2'X_e' + H_3'X_e) \tag{12G}$$

The first term $H_2'\,H_3'$ is irrevalent because function coefficients cannot operate one on another. Equations 12(F) and 12(G) specify the transition from one state into another because the signal Z_{e-1} is the input into the delay element and its state is delayed for one time increment only. Applying equations 3(A) and 3(B) to the paired equations yields eventually

$$\begin{Vmatrix} \dot{Z}_e \\ \dot{Z}_e' \end{Vmatrix} = \begin{Vmatrix} H_0 & H_1 \\ H_2' & H_3' \end{Vmatrix} \begin{Vmatrix} X_e' \\ X_e \end{Vmatrix} \tag{13}$$

This matrix is called transfer matrix and must comprise in each row of the (*h*) matrix at least a single ONE. The implication is of such significance that it is regarded as design criterion 1.

This restriction limits the number of useful functions of two variables to seven only as listed in Table I. Any other logic performed in a circuit according to Fig. 2 is either redundant or inoperative with respect to *Z*.

A sequence (*z*) may be computed with the aid of equation 13 by means of two equations only because of the single input (*x*) being either 0 or 1. The first line in Table I comprises the logic (h_2) = (0100) in the order of the function coefficients. H_2 and H_3 has to be complemented in equation 13 and it yields equations 14(A) and 14(B). Colons within brackets denote periodic sequences.

$$X = 1: \quad (\dot{z}) = \begin{Vmatrix} 0 & 1 \\ 1 & 1 \end{Vmatrix} \begin{Vmatrix} 0 \\ 1 \end{Vmatrix} = \begin{Vmatrix} 1 \\ 1 \end{Vmatrix} = \begin{Vmatrix} \dot{Z} \\ \dot{Z}' \end{Vmatrix} \tag{14A}$$

$$X = 0: \quad (\dot{z}) = \begin{Vmatrix} 0 & 1 \\ 1 & 1 \end{Vmatrix} \begin{Vmatrix} 1 \\ 0 \end{Vmatrix} = \begin{Vmatrix} 0 \\ 1 \end{Vmatrix} = \begin{Vmatrix} \dot{Z} \\ \dot{Z}' \end{Vmatrix} \tag{14B}$$

Z will be switched off for $X = 0$ and on or off for $X = 1$.

A feedback loop as shown in Fig. 3 comprising a static storage element as a flip-flop requires a set signal $\dot{Y}$ and a reset signal $\dot{Y}'$. Each signal may be generated by a separate logic *g* and *h*.

$$\dot{Y} = Y'(G_0X' + G_1X) + Y(G_2X' + G_3X) \tag{15A}$$

$$\dot{Y}' = Y'(H_0X' + H_1X) + Y(H_2X' + H_3X) \tag{15B}$$

The set signal $\dot{Y}$ is irrelevant for Y = 1 and similarly $\dot{Y}'$ for Y′ = 1. Higher reliability of the feedback loop will be assured for the condition $G_2, G_3, H_0, H_1 = 0$, which is stated as design criterion 2:

$$\begin{Vmatrix} G_2 & G_3 \\ H_0 & H_1 \end{Vmatrix} = 0 \tag{15C}$$

Similarly to equation 13, the transfer matrix for the trigger signals $\dot{Y}$ and $\dot{Y}'$ may be written

$$\begin{Vmatrix} \dot{Y} \\ \dot{Y}' \end{Vmatrix} = \begin{Vmatrix} G_0 & G_1 \\ H_2' & H_3' \end{Vmatrix} \begin{Vmatrix} X' \\ X \end{Vmatrix} \tag{16}$$

6. Transfer Function of General Networks

In the primitive case demonstrated, the transfer matrix is obtained as a 2-dimensional operator for the input variable *X*; it yields the differentials or trigger signals for a single storage element only. It has been shown, in equation 18 of reference 1, that the function matrices as such are the necessary and sufficient expressions for the behavior of a net and that the states or combinations are properly defined by the dummy index *i*. This concept gives the required flexibility to deal with nets of higher complexity.

For a general network as shown in Fig. 4 a set of independent inputs *X* and dependent inputs *Y* originated in feedback loops are considered. The conversion of a sequential input (*x*) into parallel is not significant in this context. The static logic has one set of exterior outputs *Z* and another set of interior outputs or trigger signals $\dot{Y} + \dot{Y}'$ controlling storage elements in the feedback loops. The logic (*f*) operates solely on the instantaneous combinations of all inputs *X* and *Y* to generate both sets of outputs $\dot{Y} + \dot{Y}'$ and *Z* respectively.

The dummy index *i* specifies the combination of all inputs $X + Y$, and accordingly the output, e.g.,

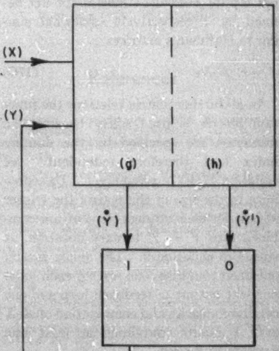


Fig. 3. General net

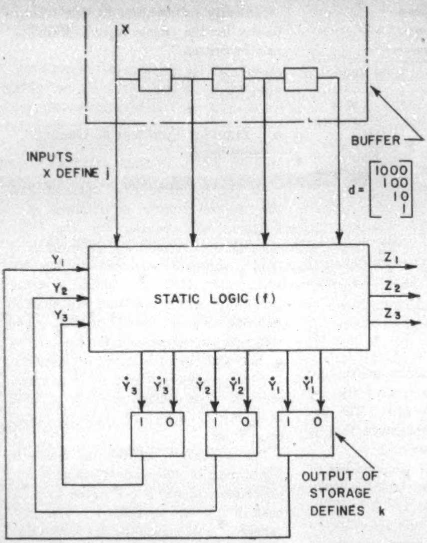


Fig. 4. Feedback with static storage element

$$Z_r = F_{ri} N_i \qquad (17A)$$

The minterm N_i may be divided into one minterm related to the independent input combination of X and another N_k for the dependent combination of Y.

$$N_i = N_j N_k \qquad (17B)$$

For convenience i may be written in such a form as to preserve the numerical terms for j, k. This allows the direct correlation to the states of the independent inputs X or the states of the net defined by Y respectively somewhat similar to Huffman's matrices.

$$Z_r = F_{rj} N_j N_k \qquad (17C)$$

In all further matrix relations the function matrix is used solely because the minterms are specified by the dummy index and therefore redundant; see equation 17(B) of reference 1. To economize in the size of the matrix the trigger signals for each storage element are combined into $\dot{Y} + \dot{Y}'$ ternary notation, as stated in equation 2. The single matrix obtained comprises one row for each independent output or feedback loop and one column for each valid combination of all X and Y inputs containing at least one nonzero coefficient.

A transfer function matrix of this type is the complete description of a sequential logic for its outputs and states depending on the imposed input conditions. The complete sequence of states and output signals may be derived from this matrix by operating on the dummy index $i = j,k$ according to each column. By writing the sequence of (i) similar to (9) the sequence of states is already specified by the section (k) of (i).

To demonstrate the method a frequently quoted problem has been chosen. It concerns the design of an indicator for the direction of rotation with a coded disk. Departing from the conventional concept of three segments, two half-circular arcs are contemplated displaced 90 degrees. From the brushes or photocells two pulse trains will be available and should cause a distinct signal Z for the direction denoted as negative.

Value of x	Positive Direction	Negative Direction
x_1	$=(:0110:)$	$=(:0110:)$
x_2	$=(:0011:)$	$=(:1100:)$

(18A)

The colon within the brackets may indicate that the pulse trains are periodic. According to the convention the least significant bits refer to an arbitrary zero-time increment. Its length depends on the speed of the disk and is irrelevant for a logic of ideal response. As an aid for discussion the individual inputs X_1, X_2 are listed below the index j.

$$i = \begin{cases} k = 0\,0\,0\,0 \;\; 1\,1\,1\,1 \\ j = 0\,1\,2\,3 \;\; 0\,1\,2\,3 \end{cases} \quad \begin{array}{c}\text{column}\\ \text{index}\end{array}$$

$$\begin{aligned} X_1 &= 0\,1\,0\,1 \;\; 0\,1\,0\,1 \\ X_2 &= 0\,0\,1\,1 \;\; 0\,0\,1\,1 \end{aligned}$$

$$Y = 0\,0\,0\,0 \;\; 1\,1\,1\,1$$

$$\left\| \begin{matrix} \dot{Y} + \dot{Y}' \\ Z \end{matrix} \right\| = \left\| \begin{matrix} 0 & 1 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{matrix} \right\| \qquad (18B)$$

One state j must be chosen to switch a storage element Y on, e.g., for $j = 1$, $\dot{Y} = 1$. Depending on the direction of the disk two sequences may be written as follows:

Positive direction:

$$(i+) = \left\| \begin{matrix} k \\ j \end{matrix} \right\| = \left\| \begin{matrix} :0\;0\;0\;0\;1\;1: \\ :0\;2\;3\;1\;1\;0: \end{matrix} \right\| \qquad (18C)$$

Negative direction:

$$(i-) = \left\| \begin{matrix} k \\ j \end{matrix} \right\| = \left\| \begin{matrix} :0\;0\;1\;1\;1\;0: \\ :0\;1\;1\;3\;2\;2: \end{matrix} \right\| \qquad (18D)$$

The behavior of the net is therefore completely described by the two rows of the transfer matrix; see equations 18(B). Redundancy elimination by means of the r-matrix (equation 9 of reference 2) is directly applicable. In conventional form the final design equations may be rewritten from the transfer matrix as follows:

$$\dot{Y} = X_1 X_2'$$

$$\dot{Y}' = X_1' X_2'$$

$$Z = Y X_2 \qquad (18E)$$

A circuit of this kind flashes a signal for the negative direction of the disk. If a permanent signal occurs the disk is at still stand with X_2 on.

7. Synthesis by Inversion

The ultimate goal of the previous analysis is a direct routine for synthesis of a network in order to produce a certain output sequence for a specified input sequence. A sequence (x) passing through a delay chain may always be substituted by the minterms (n) as presented to the logic (f). For each minterm N_{si} the output Z_s is already specified for the time increment s and may be transcribed directly. Only where contradiction exists a decision has to be made for the proper value of F_i. Expressing the process in matrix notation for the coefficients involved yields an inversion eventually. The operator : denotes an inversion. In this case it indicates that Z_s has to be referenced to N_{si} or the index (i) proper.

$$Z_s = N_{si} F_i \qquad (19A)$$

$$F_i = Z_s : N_{si} \qquad (19B)$$

$$(i) = (y)^* = [(x)(d)]^* \qquad (19C)$$

As an illustrative example a problem similar to the foregoing is used with the restriction that a single channel is available only to transmit a pulse train. The pulse train shall generate a signal for the sequence (x) and no signal for the reversed sequence $(x)^-$. According to Section 1 the sequence of minimum length must be found which is distinguishable from its reverse. Because the first and necessarily the last bit must be ONE, at least two more bits of complementary value must be inserted. Following are the desired output sequences (z):

$$(x)=(1101) \qquad (x)^-=(1011)$$

$$(z)=(11111) \qquad (z)'=(00000)$$

$$(i+)=\left\| \begin{array}{ccccccc} 1&1&0&1&&& \\ &1&1&0&1&& \\ &&1&1&0&1& \\ &&&1&1&0&1 \end{array} \right\| =(\mathit{10}, 14, \mathit{6}, 13, 5, 2, \mathit{1})$$

$$(i-)=\left\| \begin{array}{ccccccc} 1&0&1&1&&& \\ &1&0&1&1&& \\ &&1&0&1&1& \\ &&&1&0&1&1 \end{array} \right\| =(\mathit{10}, 4, 12, 15, \mathit{6}, 3, \mathit{1})$$

Comparing the sequences $(i+)$ and $(i-)$ three common terms *10*, *6*, *1* as italicized will be noted. The first and last one are irrelevant but *6* will be disallowed for (z) and therefore for this circuit the compromise for (z) = (10111) must be considered. Deriving the function matrix (f) connected to three delay elements from $(i+)$ and $(1-)$ is straightforward. Any *Fi* not relevant is marked — in order to indicate freedom in minimization.

$$(f)=(0010010-\ 0-0110-\!-)$$

Conclusions

Referring to the work by others on sequential circuits, Huffman is primarily concerned with asynchronous logic similar to relay circuits[5] while Mealy deals with synchronous systems related to electronic data processing.[4] Synchronism, literally speaking, implies coincidence in phase and pulse rate in that respect that pulse form, timing of pulse edges, and duration must conform to the standards established for the system. With increasing distances among the units of digital systems the phase condition cannot be maintained and has to be eventually abandoned. Remote inputs can be sampled locally but either the system must be able to process such data on line or buffer storage will be required. Open loop systems with delay chains are related to the broader concept of synchronous systems without phase restriction. Phase-controlled operation relieves the restriction specified in section 1.

Matrix algebra applied to the design of sequential logic has the advantage of a complete and explicit representation of a net adaptable to computer routines. The scope of the method is general enough to utilize multilevel logic and any type of operation either asynchronous or synchronized. Methods for analysis of matrix logic, the expansion by added input variables and the reduction routines derived previously, are directly applicable for sequential operations.

The examples used in the different phases were chosen for illustrative purposes only and are elementary. Actual design problems with the usual order of more than ten inputs and several feedback loops are definitely beyond the scope of this paper and depend on the available computer facilities. Desirable are binary computers with subroutines for logical operations and individual access to several 10,000 bits storage. Computers of low speed and restricted capacity may be improvised with sets of paper tape punches and readers with a separate counter for each unit recording the indexes i, j, and k. Subsequent papers are in preparation dealing with matrix routines for symmetric logic and simultaneous equations for logical functions.

Appendix. Glossary of Terms and Symbols

Symbols and notations used in this series on matrix algebra are compiled and referenced with respect to references 1 and 2, and to this paper. In the process of standardizing the symbols throughout the series some modifications were unavoidable and have been marked with an asterisk following the reference.

b=base or radix of positional notation.

(d)=delay matrix of coefficients D_e; output sequence of delay circuit during time increment e for a unit pulse as input at zero time; see equation 4.

e=time index for elapsed time increment since zero time in delay matrix (d) or transfer function; see equation 4.

(f)=function matrix of coefficients F_i defining the output of combinatorial logic for input combination N_i; see reference 1, equation 13.

$(f)'$=complement of (f); $(f)+(f)'=(1)$; $(f)^-{}'=(f')^-$; see reference 1, equation 13.

(g)=expanded function of coefficients G_i, G_j; see expansion theorem; see reference 2, equation 3.

i=dummy index denoting the coincidence condition of p binary inputs with each bit corresponding to one input; for binary inputs preferably in octal notation; see reference 1, equation 9(A).*

(i)=sequence of i; see equation 9.

j=dummy index similar to i; preferred for independent inputs; see reference 1, equation 11.

(j)=sequence of j; see equation 17(B).

k=index for states similar to i; see equation 17(B).

(k)=sequence of states; see equation 17(B).

(n)=minterm matrix for all coincidence conditions N_i of all inputs into combinatorial logic as defined by i; see reference 1, equation 8.

N_i, N_j, N_k=specific minterms; see equation 17(B).

p=number of input variables; see reference 1, equation 11; length of a sequence; see section 1.

q=time index in an input sequence; see equations 5 and 7.

r=row index; see equation 17(A).

(r)=redundancy matrix; see reference 2, equation 9.

s=time index of output sequence; mostly $s=e+q$; see equations 5 and 7.

$Y_s=F_{si}\times N_i$=matrix notation in Einstein convention; mostly dot product if not specified otherwise; see reference 1, equation 4(A).

(y)=function matrix for combinatorial output Y; see reference 1, equation 18.

$(y)^*$=transpose of (y); see equation 9.

$(y)^-$=reverse of (y); see reference 1, equation 14.*

$\dot{Y}$, $\dot{Y}'$=transition of logical function; see equation 2.

transfer function=composite matrix of trigger or transfer outputs into feedback loops and external outputs; see equations 13 and 16.

: =inversion; mostly referencing Z_s with respect to i as specified by N_{si}; see equation 19(B).

(: . . . . :)=periodic sequence; see equation 18(A).

( )=short notation for matrix in lieu of bold face type; see equation 4.

References

1. MATRIX ANALYSIS OF LOGICAL NETWORKS, E. J. Schubert. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, Mar. 1958, pp. 10–13.

2. MATRIX SYNTHESIS OF HIGH-SPEED LOGIC, E. J. Schubert. *Ibid.*, vol. 78, Mar. 1959, pp. 4–8.

3. GEDANKEN EXPERIMENTS ON SEQUENTIAL MACHINES, E. F. Moore. "Automata Studies," Princeton University Press, Princeton, N. J., 1956.

4. A METHOD FOR SYNTHESIZING SEQUENTIAL CIRCUITS, G. H. Mealy. *Bell System Technical Journal*, New York, N. Y., vol. 34, no. 5, 1955, p. 1045.

5. THE SYNTHESIS OF SEQUENTIAL SWITCHING CIRCUITS, D. A. Huffman. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 257, no. 3, Mar 1954, pp. 161–90.

6. LOGICAL CALCULUS, E. J. Schubert. *Westinghouse Electronics*, Baltimore, Md., Oct. 1956.

Simultaneous Logical Equations

Matrix Logic IV

E. J. SCHUBERT
ASSOCIATE MEMBER AIEE

Synopsis: A new method for reducing logical problems of higher complexity based on matrix logic is presented. Sets of propositions related to a set of constraints are transformed into minimum form without iterative simplifications. The savings in design effort and hardware for the eventual circuit are demonstrated in a reference problem.

THE DESIGN of a control unit of a digital computer with a minimum of hardware is a typically logical problem. The conditions to be satisfied are given in a table of control signals versus instructions. These control signals may be regarded as constraints and the instructions as propositions.

A constraint is said to be compulsory/disallowed if the proposition is affirmative/negative for this particular constraint.

In a recent publication, the fact has been contemplated that the total number of constraints be a valid measure for the complexity.[1] In that concept, propositions are divided arbitrarily into two groups, the first including propositions with a greater number of constraints and the second group including the remainder. The first group only is said to be correlated to minterms in one-to-one correspondence to constraints.

The object of this paper is to demonstrate a simpler algorithm without iterative simplifications, which may be most efficiently applied to any combinational problem of higher complexity. The matrix notation used throughout the paper has been derived in previous papers. Compared with any other notation, matrix logic offers the most explicit presentation directly correlated to complete truth tables, and it may be adapted to computer routines for problems of higher complexity.[2–4]

Merging of Sets

The design of digital machines implies the mechanization of decoders and encoders similar to control units with free choice of the instruction code. Any instruction corresponds to a proposition or a declarative statement specifying a subject in reference to a set of constraints or gating signals. Eventually the propositions have to be reduced to logical nets which perform equivalent logical functions on input signals. The advantage of matrix notation has been demonstrated as a concept for processing complete truth tables efficiently. Such truth tables define a certain coincidence condition of the input variables by the index i of a minterm N_i, and express the binary output Y_z by the dot product of coefficients F_{zi} and N_i.

Propositions Y_z may be defined with respect to the constraints Z_k by a matrix of coefficients C_{zk}. The difference between logical functions and propositions is the fact that the latter are governed by ternary sets. The purpose of the proposed method is the correlation of ternary set functions c defining the outputs Z_k with respect to the inputs Y_z.

Any problem may be stated in sets of outputs Z_k depending on a set of encoding inputs Y_z. The solution implies a set of simultaneous equations of unions.

$$Z_k = \bigcup_z C_{kz} Y_z \tag{1}$$

The merits of the design depend on the correlation between m and k or the eventual transformation T_{km} of Z_k into Z_m.

$$Z_m = \bigcup_z C_{mz} Y_z \tag{2}$$

$$C_{zm} = C_{zk} T_{km} \tag{2A}$$

$$(c_m) = (c_z)(t) \tag{2B}$$

To find a suitable transform (t), the matrix (c) must first be defined.

The outputs Z_k, Z_m are binary signals similar to the input Y_z. These logical functions Z_k must satisfy the propositions Y_z in such a way that all constraints C_{zk} are met. Frequently a number of constraints is irrelevant in some of the propositions. This irrelevance of constraints is an additional but single state over the b states the logical functions may assume. Therefore a base $b+1$ is necessary and sufficient to describe propositions in terms of C_{zk} in a system of linear forms with the use of matrix algebra.

Consistent with numbering conventions the following symbols are proposed for binary logic and the corresponding ternary constraints:

Symbol	Binary Logic	Ternary Logic
0	false	false
1	true	true
2		irrelevant

To process ternary coefficients, operators have to be defined as in binary logic. Besides the three basic operations NOT, AND, and OR, the IDENTITY operation is introduced and denoted by $\equiv$.

NOT

C	C'
0	1
1	0
2	2

(3A)

AND

AB	0	1	2
0	0	0	0
1	0	1	1
2	0	1	2

(3B)

OR

$A+B$	0	1	2
0	0	1	2
1	1	1	2
2	2	2	2

(3C)

IDENTITY

$A-B$	0	1	2
0	1	0	1
1	0	1	1
2	1	1	1

(3D)

For any contradiction the identity function is zero, $A \equiv B = 0$.

Theorem I: Constraints common to several propositions may be merged into a single constraint, providing there is identity in the individual terms.

An output Z_k depending simultaneously on several propositions Y_z may be regarded as a vector with the components C_{zk}. Because of the peculiar feature of irrelevance any component denoted by the symbol 2 is not defined and constitutes a degree of freedom. Those components with values 0 or 1 are fixed. Any two vectors may be merged into a single one as long as the components in k are either identical or at least one of the components has the value 2.

The convention is established that the matrix is written with the column index k numbering the constraints and the row index z denoting the individual propositions. This z-by-k matrix of coefficients

---

Paper 59-1192, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted June 11, 1959; made available for printing August 10, 1959.

E. J. SCHUBERT, formerly with the Burroughs Corporation, Paoli, Pa., is now with Monitor Systems, Inc., Fort Washington, Pa., subsidiary of Epsco, Inc.

Table I. Design Problem of a Control Unit

Instructions Y_z (propositions)	Gate Signals $Z_k = \bigcup_z C_{kz} \cdot Y_z$ (constraints)																			
k=	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
z= 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	1	2	1	1	1	1	1	1	1	1	1	1	2	2	2	0	0	1
4	0	2	1	0	1	2	2	1	2	0	2	1	2	0	0	1	0	0	1	2
5	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0
6	0	1	2	2	2	2	2	2	2	2	0	0	0	2	2	2	2	0	1	2
7	2	2	2	2	1	2	2	2	1	0	1	1	2	2	2	2	1	2	1	2
8	2	2	2	2	2	2	0	2	2	2	2	2	0	0	0	2	1	2	2	2
9	2	2	0	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	1
10	2	2	2	2	2	2	2	2	1	2	2	2	2	2	0	2	2	1	2	2

"0"...OFF; "1"...ON; "2"...irrelevant

Table II. Merging of Columns $C_{kz} \rightarrow C_{mz}$

m	1	2	3	4	$\underline{3}$	$\underline{3}$	$\underline{3}$	$\underline{3}$	$\underline{3}$	$\underline{4}$	$\underline{3}$	$\underline{3}$	$\underline{4}$	$\underline{4}$	$\underline{4}$	$\underline{3}$	5	$\underline{5}$	6	7
k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
z= 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	1	$\underline{1}$	1	1	1	1	1	1	1	1	1	1	.	.	$\underline{0}$	0	0	1
4	0	.	1	0	1	.	.	1	.	0	.	1	.	0	0	1	0	0	1	.
5	.	0	$\underline{0}$	0	.	0	.	0	.	0	.	0	.	0	.	0	$\underline{0}$	0	.	0
6	0	1	$\underline{0}$	$\underline{0}$	.	.	.	.	.	.	0	0	0	.	.	.	$\underline{0}$	0	1	.
7	.	.	$\underline{1}$	$\underline{0}$	1	.	.	.	1	0	1	1	.	.	.	.	1	.	1	.
8	.	.	$\underline{0}$	$\underline{0}$	.	.	0	.	.	.	.	.	0	0	0	.	1	.	.	.
9	.	.	0	.	.	.	.	.	.	.	0	.	.	.	.	.	.	.	.	1
10	.	.	$\underline{1}$	$\underline{0}$	.	.	.	.	1	.	.	.	.	.	0	.	$\underline{1}$	1	.	.

"2"...deleted (blank) $\underline{0}, \underline{1}$...entries due to merging

C_{zk} of radix $b+1$ is called the (*c*) matrix of constraints.

Theorem I implies that any two columns of (*c*) may be merged providing that all elements in the columns satisfy equation 3(D) or

$$C_k \equiv C_{k+j} = 1 \qquad j \geq 1 \tag{4A}$$

Therefore the shifted index $k+j$ is discarded and the merged coefficients $\bar{C}_k$ are governed by ternary operations:

$$\bar{C}_k = (C_k \equiv C_{k+j}) \bigcap_z C_{kz} C_{(k+j)z} \tag{4B}$$

This reduced matrix should be reordered from C_{kz} into C_{mz}.

Similarly one instruction might be the subset of another, and therefore superfluous. Accordingly the matrix C_{mz} must be transposed and processed in the same way towards C_{rm}; r is the new order superseding z. The restored matrix C_{mr} is a reduced form and should comprise a lesser number of coefficients than the original matrix C_{kz}.

Nonredundant Encoding

To establish an algorithm for an optimum binary code of Y_z requires a measure for the significance of useful coefficients C_{mz} or C_{kz} respectively. Eventually the logical functions $Y_d = Y_d(X_d)$ are found which provide a unique code.

Theorem II: A nonredundant code for redundant words comprises the bits with the highest entropy only.

The entropy H_k describes the probability distribution of ZERO's and ONE's in the set C_{zk} for i_k ONE's in the total number of n words.

$$H_k = p_k \log p_k + (1-p_k) \log (1-p_k); \quad p_k = i_k/n \tag{5A}$$

This function has a maximum at $p = 0.5$:

$$dH = 0 = \log p - \log (1-p)$$
$$p/(1-p) = 1; \; p = 0.5 \tag{5B}$$

A more convenient relation with a maximum of the same value is

$$p_k(1-p) = i_k(n-i_k)/n^2 \tag{5C}$$

Because n is constant, the numerator suffices:

$$D_k = i_k(n-i_k) \tag{6}$$

Theorem III: The sole measure for the significance D_k for decoding purposes, called rank of a bit k, is the number i_k of ZERO's or ONE's in bit k, whichever is less. Bits of rank zero are trivial.

In a function (*f*) the number of nonzero terms may be denoted by $|F|$. If (*f*) depends on p variables then the total number of terms either true or false is given as

$$|F| + |F'| = 2^p \tag{7}$$

where $|F|$ is the number of coefficients $F_i = 1$.

Any function requires for its realization not more minterms/maxterms than the number of ONE's/ZERO's.[3] Accordingly, the number of required terms eventually depends on the lesser of $|F|$ or $|F'|$. The statement of equation 7 is true for the whole system and indistinguishable for any function. Therefore the sum of constraints in any proposition is irrelevant for any classification of one proposition with respect to another.

An order d of the bits k (or m) can be established corresponding to the rank D_k in such a way that the bits are sorted properly as to their significance for decoding. Most significant bits may be placed to the leftmost side of the words; the least significant bits are expected to be irrelevant or redundant.

Theorem IV: Any set of coded words is uniquely distinguishable, if discrimination within the pairs of "most similar" words is possible. Two words are most similar if they are adjacent in numerical order, and if they are considered as numbers written in order of rank D_k.

Sufficient for proper discrimination is a difference in at least a single bit of two most similar words. If there is a difference in more than one bit, the discrimination in the bit of highest rank is preferable. This rule will assure that bits with highest probability or significance will be considered for the nonredundant code of Y_z or Y_r.

Optimization of Implicit Codes

A set of words with some unspecified bits may be designated as an "implicit code." This code is ternary according to matrices 3(A), (B), (C), and (D). Discrimination of two such code words is possible only if, for at least one bit k, the identity function 3(D) vanishes.

Table III. Merging of Rows $C_{mz} \rightarrow C_{mr}$

m =	1	2	3	4	5	6	7	
z= 1	1	1	1	1	1	0	0	r= 1
2	1	1	0	0	0	0	0	2
3	1	0	1	1	0	0	1	3
4	0	.	1	0	0	1	.	4
5	.	0	0	0	0	.	0	5
6	0	1	0	0	0	1	$\underline{1}$	6
7	.	.	1	0	1	1	.	7
8	.	.	0	0	1	.	.	8
9	.	.	0	.	.	.	1	$\underline{6}$
10	.	.	1	0	1	.	.	$\underline{7}$

Table IV. Order d According to Significance

m=	1	2	3	4	5	6	7
r= 1	1	1	1	1	1	0	0
2	1	1	0	0	0	0	0
3	1	0	1	1	0	0	1
4	0	.	1	0	0	1	.
5	.	0	0	0	0	.	0
6	0	1	0	0	0	1	1
7	.	.	1	0	1	1	.
8	.	.	0	0	1	.	.
i_m	3	3	4	2	3	3	2
0_m	2	2	4	6	5	3	3
t_m	3	3	-	-	-	2	3
D_m			1	1	1		
	*	*	6	2	5	*	*
d	2	3	1	7	6	4	5

* maximum value for D_m possible

$$\bigcap_k C_{kz} \equiv C_{k(z+j)} = 0 \tag{8}$$

A quantitative discrimination depends on the algebraic relation

$$\sum_k 1 - C_{kz} \equiv C_{k(z+j)}) = E_{z:z+j} \tag{9}$$

Providing one of the words C_z comprises several irrelevant bits $C_{kz}=2$, E may be increased by replacing in these bits 2 by 0 or 1 according to equation 10.

$$\bar{C}_{kz} = (C_{kz} \equiv C_{k(z+j)})(1 - C_{k(z+j)}) \tag{10}$$

Theorem V: In an implicit code with some bits unspecified in some words, a bit k must be considered as possibly of highest rank if neither the number of ZERO's nor ONE's exceeds half of the number of words.

The significance of any bit k was previously defined by D_k in equation 6. If a bit or column k comprises at least irrelevant entries 2, then the upper bound for D_k is

$$\text{for } t \geq n/2 - i_z > 0;\ D_k \leq n^2/4 \tag{11}$$

Depending on system requirements for the final code, any 2 may be replaced by 0 or 1 to improve discrimination in equation 9. It is therefore possible that the eventual distribution of 0's and 1's yields the maximum D_k. In some cases, where the instruction code must operate the gates directly without any auxiliary control units, all the bits must appear in the instructions.

Table V. Sorting in Similarity

d	1	2	3	4	5	6	7	
s= 7	1	1	1	0	0	1	1	r= 1
4	0	1	1	0	0	0	0	2
6	1	1	0	0	1	0	1	3
5	1	0	.	1	.	0	0	4
1	0	.	0	.	0	0	0	5
3	0	0	1	1	1	0	0	6
8	1	.	.	1	.	1	0	7
2	0	.	.	.	.	1	0	8

Application to Design of Control Unit

Based on theorems I through V, the following steps are suggested for the solution of logical design problems involving simultaneous propositions related to a larger number of constraints. The prime objective is to establish and maintain consistent orders throughout all steps.

1. Formulate the propositions Y_z in matrix coefficients C_{zk} with respect to the constraints Z_k.
2. Merge constraints by the operation described in equation 4(A) and (B), transforming C_{zk} into C_{rm}, correlating r to z and m to k.
3. Determine ranks of variables X_d by equation 6, correlating d to m to k.
4. Reorder C_{rm} into C_{rd} and sort as sequence of ternary numbers.
5. Determine difference of adjacent words in the most significant bit d only.
6. Retain for code of Y_z or $Y_r = Y(X_d)$ only those bits d in which a difference occurs.
7. The remaining outputs Z_m must be formed as logical function $Z_m = Z_m(X_d)$.

To evaluate the merits of this approach, a design problem is taken from a recent publication and described in Table I.[1] Ten instructions Y_z ($z=1, \ldots 10$) are given with respect to 20 gating signals Z_k ($k=1, \ldots 20$). Irrelevant conditions are entered as 2.

In Table II all columns are evaluated taking two columns at a time with respect to identity 4(A). Starting from the left, the columns $k=1$, 2 remain isolated. Column $k=3$ can be merged with all other columns marked $m=\bar{3}$. Merged coefficients according to equation 4(B) have to be entered accordingly. This procedure reduces the 20 original columns in order k to 7 columns on order m. This process is repeated for the rows in Table III and eliminates $z=9$, 10 by $r=\bar{6}$, $\bar{7}$ respectively.

To determine the significance of the bits m for decoding purposes a count of 0's and 1's is made in Table IV. Whenever theorem IV applies, an asterisk is inserted for D_m. The rank is the highest for $D_3=16$ ($m=3$), and accordingly assigned to $d=1$. Those bits marked by an asterisk are possibly of highest rank also, and are therefore placed in consecutive order $d=2, \ldots 5$.

Table VI. Minimum Code

	INPUT			OUTPUT			
d	1	2	3	4	5	6	7
s= 1	0	0	0	.	0	0	0
2	0	$\bar{1}$	0	.	.	1	0
3	0	$\bar{0}$	$\bar{1}$	1	1	0	0
4	0	1	1	0	0	0	0
5	1	0	0	1	.	0	0
6	1	1	$\bar{0}$	1	.	0	0
7	1	1	1	0	0	1	1
8	1	$\underline{0}$	$\underline{1}$	1	.	1	0
m	3	1	2	6	7	5	4
k				1	2	1	
	3	1	2	9	0	7	4
				Z_m			

After reordering in d corresponding to their ranks in Table V, the words are sorted according to their similarity. Index $s=1$ designates the word $r=5$ as the lowest binary number.

The final matrix C_{ds} may be adapted to comply with system specification for the input code proper. For a minimum code for eight words, with three bits $d=1$, 2, 3, some additional entries $\bar{0}$, $\bar{1}$ are essential, as shown in Table VI. The inputs will operate the output signals $m=1, 2, 3$ directly. Remaining outputs must be formed in an encoder matrix.

If systems specifications demand a given code, e.g., Teletype for the instructions, the first five bits $d=1, \ldots 5$ may be matched with appropriate Teletype symbols. Then, two additional outputs $m=4$

Table VII. Matched Code

(Five-Bit Teletype Code)

	INPUT					OUTPUT		TELETYPE CODE
d	1	2	3	4	5	6	7	
s= 1	0	1	0	1	0	0	0	R
2	0	$\bar{0}$	1	$\bar{1}$	0	1	0	N
3	0	$\bar{0}$	$\bar{1}$	$\bar{1}$	$\bar{1}$	0	0	M
4	0	1	1	0	0	0	0	I
5	1	0	0	1	0	0	0	D
6	1	1	$\bar{0}$	0	$\bar{1}$	0	1	W
7	1	1	1	0	0	1	1	U
8	1	$\underline{1}$	$\underline{1}$	1	$\underline{0}$	1	0	K
m	3	1	2	6	7	5	4	
k				1	2	1		
	3	1	2	9	0	7	4	
						Z_m		

Vol.
Pt. 1
1959

5 are required only, because the inputs are identical with five outputs, as illustrated by Table VII.

Conclusions

Drastic reductions in simultaneous logical equations can be achieved by merging constraints without contradiction. These reduced forms may be coded efficiently, depending on the symbol distributions within the code bits. Direct applications of this approach are the design of control units or code translators.

The explicit algorithms are easily adaptable to computer programs for logical design routines with a modest demand on memory capacity.

Further improvements of this approach are under investigation, particularly to establish a unique order in k and z prior to merging. These orders are related to the distribution of the coefficients C_{zk}.

References

1. SIMPLIFICATION OF A CLASS OF BOOLEAN FUNCTIONS, E. Hirschhorn. Association for Computing Machinery, New York, N. Y., vol. 5, no. 1, 1958, p. 67.

2. MATRIX ANALYSIS OF LOGICAL NETWORKS, E. J. Schubert. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, Mar. 1958, pp. 10–12.

3. MATRIX SYNTHESIS OF HIGH SPEED LOGIC, E. J. Schubert. *Ibid.*, vol. 78, Mar. 1959, pp. 4–8.

4. MATRIX ALGEBRA OF SEQUENTIAL LOGIC—MATRIX LOGIC III, E. J. Schubert. *Ibid.*, see pp. 1074–79 of this issue

5. SYMMETRIC SWITCHING FUNCTIONS—MATRIX LOGIC V, E. J. Schubert. *Ibid.*, pp. 1083–87.

Symmetric Switching Functions

Matrix Logic V

E. J. SCHUBERT
ASSOCIATE MEMBER AIEE

Synopsis: Symmetric logical functions are classified in ordinary and threshold function. The first class is defined as being operative for only k of p inputs being active. Threshold functions are operative for at least k of p inputs. Ordinary symmetric and threshold functions are mutually related and one class may generate functions of the other class. Transistor circuits are already extensively used to perform threshold functions in form of resistor-coupled nonlinear amplifiers. Recently a new logical element has been conceived which performs ordinary symmetric logic. To evaluate the merits of these logical elements efficiently a method has been developed which may be adapted to logical design by computer routines. Solution of problems involving up to ten variables is feasible without the aid of computers.

COMBINATIONS of inputs which are irrespective of any preference or relative position of one input to another are commonly known as symmetric switching functions. A typical logic of this type is an output depending on any two out of five inputs.

The significance of these logical functions was due to the economy in the design of relay networks using transfer contacts.[1] High-speed logic based on solid state components has no direct equivalent to the transfer contact and therefore the concept of symmetric functions has been neglected.

Ordinary Symmetric Function

A symmetric function depending on k and only k active inputs, out of the total number of p inputs, may be called ordinary, and corresponds to the combinations of p different things, all different, taken k at a time:

$$_pC_k=\binom{p}{k}=\binom{p}{p-k} \tag{1}$$

Conventional logic elements are described by the primitive relations for which one and only one combination is unique. The mathematical interpretation for C is correct for both extremes $k=0$ and $k=p$ respectively. The first case describes the inhibited buffer, the second case the operative gate.

In the canonical form any function Y_z may be represented by the sum of minterms or the product of maxterms. A minterm N/maxterm M is any one of the 2^p distinct AND/OR conditions of all p variables X, in which these variables are or are not complemented in any of the 2^p possible ways.[2] The suffix i of N or M is the octal equivalent of the binary number formed by writing ZERO for each complemented variable and ONE for each uncomplemented variable, in the order of their appearance.

In the preferred matrix presentation functions are sums of minterms.[3] A binary coefficient F_i indicates if the minterm N_i is applicable for the function Y_z, or expressed in matrix notation:

$$Y_z=F_{zi}N_i \qquad (y)=(f)(n) \tag{2}$$

To recognize an ordinary symmetric function in this form depends on two conditions to be met by the suffixes i:

Rule 1: The number of minterms involved must be equal to a binomial coefficient as defined in equation 1.

Rule 2: In this set of minterms one and the same input must occur as often as any other input.[4] The subset of minterms comprising one and the same input are the combinations of $p-1$ things, taken $k-1$ at a time, or

$$_{p-1}C_{k-1}=\binom{p-1}{k-1} \tag{3A}$$

As a typical example a function of five variables is given in the form

$$Y=N_{17}+N_{27}+N_{33}+N_{35}+N_{36} \tag{3B}$$

For convenience the suffixes are written in octal notation and easily converted into binary:

$$\begin{array}{ll} 17\ (\text{o}) & 01\ 111 \\ 27\ (\text{o}) & 10\ 111 \\ 33\ (\text{o}) & 11\ 011 \\ 35\ (\text{o}) & 11\ 101 \\ 36\ (\text{o}) & 11\ 110 \end{array} \tag{3C}$$

Only two binomial coefficients of a set, $p=5$, are equal to the number 5 corresponding to the number of terms:

$$\binom{5}{1}=\binom{5}{4}=5 \tag{3D}$$

The occurrence of active inputs is obtained as the sum, digit 4 in rule 2, from the foregoing tabulation of the binary suffixes. Only the second binomial coefficient satisfies equation 3(A):

$$\binom{5-1}{4-1}=\binom{4}{3}=4 \tag{3E}$$

Accordingly it can be stated that the function describes an output being operative for four out of five inputs. To express this relation in short form the bino-

---

Paper 59-205, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 20, 1958; made available for printing July 3, 1959.

E. J. SCHUBERT, formerly with the Burroughs Corporation, Paoli, Pa., is now with Monitor Systems, Inc., Fort Washington, Pa., subsidiary of Epsco, Inc.

Table I. Binomial Coefficient $\binom{p}{k}$

k =		0	1	2	3	4	5	6	7	8	9	10
p =	1	2	1									
	2	4	3	1								
	3	8	7	4	1							
	4	16	15	11	5	1						
	5	32	31	26	16	6	1					
	6	64	63	57	42	22	7	1				
	7	128	127	120	99	64	29	8	1			
	8	256	255	247	219	163	93	37	9	1		
	9	512	511	502	466	382	256	130	46	10	1	
	10	1024	1023	1013	968	848	638	386	176	56	11	1

Table II. Sum of Binomial Coefficients $\binom{p}{\geq k}=\sum_{k\leq q\leq p}\binom{p}{q}$

k =		0	1	2	3	4	5	6	7	8	9	10
p =	1	1	1									
	2	1	2	1					(Pascal's triangle)			
	3	1	3	3	1							
	4	1	4	6	4	1						
	5	1	5	10	10	5	1					
	6	1	6	15	20	15	6	1				
	7	1	7	21	35	35	21	7	1			
	8	1	8	28	56	70	56	28	8	1		
	9	1	9	36	84	126	126	84	36	9	1	
	10	1	10	45	120	210	252	210	120	45	10	1

mial coefficient is suggested to be used as a function symbol:

$$Y=\binom{p}{k}X \qquad (3F)$$

The complexity of a symmetric function is characterized primarily by the term k. Referring to a symmetric function by its degree specifies the term k irrespective of p.

Threshold Function

A function describing an output being operative for any combination of at least k out of p active inputs is called a threshold function. It is a composite function comprising a number of ordinary symmetric functions. A corresponding binomial coefficient yields

$$T=\binom{p}{\geq k}X=\binom{p}{k}X+\binom{p}{k+1}X+\ldots+\binom{p}{p}X \qquad (4)$$

Solid-state circuitry lends itself to threshold functions. To form an ordinary symmetric function by means of threshold functions, equation 4 is transformed as follows:

$$\binom{p}{\geq k}X=\binom{p}{k}X+\binom{p}{\geq k+1}X \qquad (5A)$$

Substituting for the threshold functions T_1, T_2 respectively, and S for the ordinary symmetric function, yields

$$T_1=S+T_2 \qquad (5B)$$

$$T_1T_2'=ST_2'+T_2T_2' \qquad T_2T_2'=0$$

$$S=T_1T_2' \qquad (5C)$$

After manipulating on equation 5(B) and substituting in equation 5(C) the equation for the ordinary symmetric function generated by threshold functions is obtained:

$$\binom{p}{k}X=\binom{p}{\geq k}X\binom{p}{\geq k+1}'X \qquad (5D)$$

Equation 5(D) describes the coincidence of a threshold function of at least k of p inputs with the complemented threshold function of at least $k+1$ of p inputs to produce the ordinary symmetric function of k out of p inputs.

Matrix Algebra

Any function of p variables may be expressed by equation 1 wherein i is within the bounds of 0 and $2p-1$. The principles of matrix algebra modified for logic have been published previously and are therefore considered as known for the derivation to follow.[3, 5]

A symmetric function may be written in matrix notation:

$$\binom{p}{k}X=F_{ki}N_i \qquad (6)$$

$$F_{ki}=1 \qquad \text{if } k=0, \qquad \text{then } i=0 \qquad (6A)$$

$$\text{if } k\neq 0, \quad \text{then } i_q=\sum_q 2^{n}q \qquad (6B)$$

$$1\leq q\leq {}_pC_k \qquad (6C)$$

$$0\leq n_q\leq p-1 \qquad (6D)$$

According to the definition of the symmetric function the suffixes i are the binary numbers obtained by combining k ONE's in a p bit word. Table I contains the binomial coefficients for $p\leq 10$ to determine the number of terms and/or suffixes to be calculated.

A threshold function differs in its matrix in the equation for the suffix only according to the definition that the binary number is necessarily the combination of at least k ONE's in a p bit word.

$$1\leq q\leq \sum_r {}_pC_r \qquad (7A)$$

$$k\leq r\leq p \qquad (7B)$$

The conditions for unequal exponents n_q and their bounds are as in equation 6. In Table II the sum of the binomial coefficients similar to 4 is listed as a check for the total number of terms and/or suffixes involved.

In Table III the matrices for a set of four variables are listed. The suffixes are written in octal notation, ZERO's are deleted whenever possible. Lines 1 through 4 contain the states of the variables. With respect to the ordinary symmetric function in line 8, the function coefficient F_{ki} must be ONE for any two but only two input variables being ONE.

NEGATION

A proposition being true for any combination of p things taken k at a time may still be satisfied for another proposition of $p-k$ things taken at the same time. Coincidence is definitely excluded for one proposition being true for p things taken k at a time and another proposition taking $p-k+1$ things at the same time. This logical statement is directly verified by "reversed" function matrices and may be stated in terms of threshold functions.[3]

Vol.
Pt. 1
1959

$$\binom{p}{\geq k}' X_p = \binom{p}{\geq p-k+1}^{-} X_p$$

$$= \binom{p}{\geq p-k+1} X'_p \quad (8)$$

The significance of this relation is in the equivalence of two threshold functions of different degree. In another paper the need for symmetric functions of the lowest degree has been demonstrated.[6]

In Table I, line 13, is the threshold function for at least two of four active inputs X. Its complement is line 16, and it is the reversed function of at least three of four active inputs (line 12) or equivalent to the threshold function of at least three of four complemented inputs.

$$\binom{4}{\geq 2}(X_4X_3X_2X_1) = \binom{4}{\geq 3}^{-}(X_4X_3X_2X_1)$$

$$= \binom{4}{\geq 3}(X_4'X_3'X_2'X_1') \quad (9)$$

AND Operation

Two ordinary symmetric functions of the same set but of different degree are disjoint because any minterm can be a combination of only k of p things at a time:

$$\binom{p}{k}X_p\binom{p}{m}X_p = 0 \quad k \neq m; \quad 0\ldots\text{false} \quad (10)$$

On the other hand, any threshold function gated against an ordinary symmetric function yields the latter only if the first one is of higher degree:

$$\binom{p}{\geq k}X_p\binom{p}{m}X_p = \binom{p}{m}X_p \quad k \geq m \quad (11)$$

This relation is directly obtained from equation 4.

OR Operation

Because ordinary symmetric functions of the same set but of different degree are disjoint, the union is nonredundant. The test of the suffixes of such a union of two ordinary symmetric functions would satisfy rule 1 but not rule 2.

Normalized Symmetric Matrix

Any symmetric switching circuits designed for k of p inputs will perform properly for the specified number of k inputs irrespective of p. Accordingly the number k may be mapped into the minterms designated by the dummy index i similar to equation 1.

Table III. Symmetric Function Matrices (F_{ki})

Line																			
1	Variables		X_1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
2			X_2	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
3			X_3	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
4			X_4	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
5	Suffix octal		i	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
				0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
6	$\binom{4}{0}$	C =	1	(1															0)
7	$\binom{4}{1}$		4	(0	1	1		1				1							0)
8	$\binom{4}{2}$		6	(0			1		1	1			1	1		1			0)
9	$\binom{4}{3}$		4	(0							1				1		1	1	0)
10	$\binom{4}{4}$		1	(0															1)
11	$\binom{4}{\geq 4}$		1	(0															1)
12	$\binom{4}{\geq 3}$		5	(0							1				1		1	1	1)
13	$\binom{4}{\geq 2}$		11	(0	0	0	1	0	1	1	1	0	1	1	1	1	1	1	1)
14	$\binom{4}{\geq 1}$		15	(0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1)
15	$\binom{4}{\geq 0}$		16	(1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1)
16	$\binom{4}{\geq 2}'$			(1	1	1		1				1							)

To refer to a particular case of $p=4$, the function matrices for the different values of k may be derived immediately from the frequency of 1's in the binary dummy index i.

$$\begin{array}{lcccccccccccccccccl}
i & 0&0&0&0 & 0&0&0&0 & 1&1&1&1 & 1&1&1&1 & \\
 & 0&0&0&0 & 1&1&1&1 & 0&0&0&0 & 1&1&1&1 & \text{binary} \\
 & 0&0&1&1 & 0&0&1&1 & 0&0&1&1 & 0&0&1&1 & \\
 & 0&1&0&1 & 0&1&0&1 & 0&1&0&1 & 0&1&0&1 & \\
i & 0&0&0&0 & 0&0&0&0 & 1&1&1&1 & 1&1&1&1 & \text{octal} \\
 & 0&1&2&3 & 4&5&6&7 & 0&1&2&3 & 4&5&6&7 & \\
k=0 & 1& & & & & & & & & & & & & & & & \\
k=1 & &1&1& & 1& & & & 1& & & & & & & & \\
k=2 & & & &1 & &1&1& & &1&1& & 1& & & & \\
k=3 & & & & & & & &1 & & & &1 & &1&1& & (12) \\
k=4 & & & & & & & & & & & & & & & &1 &
\end{array}$$

$$\|P_{ki}\| = (0112 \;\; 1223 \;\; 1223 \;\; 2334)$$

It has been shown previously that symmetric functions of different degree are disjoint or that each minterm occurs in a single symmetric function only. Instead of writing five matrices for different k's a single matrix, $\|P_{ki}\|$ may be formed by designating each minterm with the appropriate value of k. A significant feature of this matrix $\|P_{ki}\|$ is its symmetry if written in quadratic form.

$$|P_{ki}| = \left\|\begin{matrix} 0 & 1 & 1 & 2 \\ 1 & 2 & 2 & 3 \\ 1 & 2 & 2 & 3 \\ 2 & 3 & 3 & 4 \end{matrix}\right\| \quad \text{normalized symmetric matrix} \quad (13)$$

The $\|P_{ki}\|$ matrix has been developed for $k=10$ and is shown in Table IV. It is essentially a truth table of symmetric functions with $p=10$, independent variables replacing the true value ONE for the appropriate value of k. The reference frame in i_x and i_y is in octal notation wherein the two pairs of octal digits are interchangeable because of the symmetric feature of the $\|P_{ki}\|$ matrix.

Expansion of Symmetric Functions by Cascading

Two logical blocks are considered, the first with p inputs and a function (f), the second one with $q+1$ inputs and a function (h). The output (f) of the first block is connected to one input of the following block. The output (g) of the second block must be a Boolean function of $p+q$ variables.

A cascade of two symmetric functions implies that an output (g) from the second block will be possible under following conditions only:

$$\text{for } \binom{p}{k_1} = 0 \qquad \binom{q}{k_2} = 1$$

$$\text{number of active inputs: } k_2 \quad (14A)$$

$$\text{for } \binom{p}{k_1} = 1 \qquad \binom{q}{k_2-1} = 1$$

$$\text{active inputs: } k_2 - 1 + k_1 \quad (14B)$$

$$(g) = \binom{p}{k_1}'\binom{q}{k_2} + \binom{p}{k_1}\binom{q}{k_2-1} \quad (14C)$$

In the first case, equation 14(A), the number of active inputs required is k_2; in the second case, 14(B), the output from the first block is operative with k_1 and therefore k_2-1 active inputs into the second block are permissible. The resulting output (g) does not represent a symmetric function any more.

It has been shown that a basic matrix relation called the expansion theorem exists for any cascade of logical functions.[5]

$$(g)=(h)\left\|\begin{matrix} f' \\ f \end{matrix}\right\| \tag{15A}$$

In a particular example two symmetric functions (f), (h), may be cascaded by using the relation 15(A).

$$(f)=\binom{3}{2}=(00010110)$$

$$(h)=\binom{3}{1}=(01101000)$$

$$(g)=\left\|\begin{matrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{matrix}\right\| \left\|\begin{matrix} 11101001 \\ 00010110 \end{matrix}\right\| = \left\|\begin{matrix} 00010110 \\ 11101001 \\ 11101001 \\ 11101001 \end{matrix}\right\| \tag{15B}$$

The cascaded output (g) is an unsymmetric function as predicted.

Reduction of a Function Matrix by Symmetric Logic

Function matrices are complete truth tables in an orderly notation. Routines for processing, analyzing, and generating such function matrices have been derived elsewhere.[7, 8] Any function matrix (f) is the sum of other matrices (s), (t), (u); wherein (s) represents ordinary symmetric functions, (t) threshold functions, and (u) the remaining unsymmetric functions.

$$(f)=(s)+(t)+(u) \tag{16}$$

To extract the part $(s)+(t)$ from (f) all coefficients F_{zi} must be inspected and compared with the normalized symmetric function. If for a certain value of k all or most of its coefficients $P_{ki}=k$ coincide with some F_{zi}, then this part of (f) may be efficiently covered by a symmetric function $\binom{p}{k}$. Equations 4 and 5(A) among threshold and ordinary symmetric functions should be considered if applicable. A typical problem for six variables is solved in the Appendix.

As repeatedly shown, the function matrix should be specified to the extent only as 0's and 1's are significant. Irrelevant combinations of i should therefore be designated requiring a ternary notation for (f). In the Appendix irrelevant states are denoted by x. (For an exact treatment of binary function matrices and ternary propositions see reference 8. In this context the loosely defined ternary notation is used only to explain reduction methods.)

The dot product of $(f)\ \|P_{ki}\|$ is governed by the following relation:

$$\frac{F_i=0,1}{(f)\|P_{ki}\|=\cancel{k},k}$$

Conclusions

An extensive study of symmetric functions reveals their merits for solid-state logic and in particular for logical elements performing threshold functions. Cascaded symmetric switching circuits do not perform symmetric functions.

The generalized method is applicable to large-function matrices and permits, eventually, the evaluation of symmetric functions for problems of up to ten variables without the aid of computers. Computer routines for automation of logical design are the subject of continuing investigations and will include programs for the reduction of function matrices by means of symmetric logic.

Table IV. (P_{ki}) Matrix for $k=10$

ix	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3
iy	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
0 0	0	1	1	2	1	2	2	3	1	2	2	3	2	3	3	4	1	2	2	3	2	3	3	4	2	3	3	4	3	4	4	5
0 1	1	2	2	3	2	3	3	4	2	3	3	4	3	4	4	5	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6
0 2	1	2	2	3	2	3	3	4	2	3	3	4	3	4	4	5	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6
0 3	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7
0 4	1	2	2	3	2	3	3	4	2	3	3	4	3	4	4	5	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6
0 5	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7
0 6	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7
0 7	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7	4	5	5	6	5	6	6	7	5	6	6	7	6	7	7	8
1 0	1	2	2	3	2	3	3	4	2	3	3	4	3	4	4	5	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6
1 1	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7
1 2	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6	3	4	4	5	4	5	5	6	4	5	5	6	5	5	5	7
1 3	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7	4	5	5	6	5	6	6	7	5	6	6	7	6	7	7	8
1 4	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7
1 5	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7	4	5	5	6	5	6	6	7	5	6	6	7	6	7	7	8
1 6	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7	4	5	5	6	5	6	6	7	5	6	6	7	6	7	7	8
1 7	4	5	5	6	5	6	6	7	5	6	6	7	6	7	7	8	5	6	6	7	6	7	7	8	6	7	7	8	7	8	8	9
2 0	1	2	2	3	2	3	3	4	2	3	3	4	3	4	4	5	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6
2 1	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7
2 2	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7
2 3	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7	4	5	5	6	5	6	6	7	5	6	6	7	6	7	7	8
2 4	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7
2 5	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7	4	5	5	6	5	6	6	7	5	6	6	7	6	7	7	8
2 6	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7	4	5	5	6	5	6	6	7	5	6	6	7	6	7	7	8
2 7	4	5	5	6	5	6	6	7	5	6	6	7	6	7	7	8	5	6	6	7	6	7	7	8	6	7	7	8	7	8	8	9
3 0	2	3	3	4	3	4	4	5	3	4	4	5	4	5	5	6	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7
3 1	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7	4	5	5	6	5	6	6	7	5	6	6	7	6	7	7	8
3 2	3	4	4	5	4	5	5	6	4	5	5	6	5	6	6	7	4	5	5	6	5	6	6	7	5	6	6	7	6	7	7	8
3 3	4	5	5	6	5	6	6	7	5	6	6	7	6	6	7	8	5	6	6	7	6	6	7	8	6	7	7	8	7	8	8	9
3 4	3	4	4	5	6	5	5	6	4	5	5	6	5	6	6	7	4	5	5	6	5	6	6	7	5	6	6	7	6	7	7	8
3 5	4	5	5	6	7	6	6	7	5	6	6	7	6	6	7	8	5	6	6	7	6	6	7	8	6	7	7	8	7	8	8	9
3 6	4	5	5	6	7	6	6	7	5	6	6	7	6	6	7	8	5	6	6	7	6	6	7	8	6	7	7	8	7	8	8	9
3 7	5	6	6	7	8	7	7	8	6	7	7	8	7	7	8	9	6	7	7	8	7	7	8	9	7	8	8	9	8	9	9	10

Appendix. A Design Problem

A logical function (f) of six variables is given. x in the matrix denotes irrelevant combinations which may be either 0 or 1. Because there are fewer 1's than 0's the coefficients being 1 are multiplied with the normalized symmetric matrix $\|P_{ki}\|$.[3]

$$\begin{array}{cc} & \begin{matrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\ (f)= & \left\|\begin{matrix} 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ x & 1 & 1 & 0 & x & x & 0 & x \\ 0 & 1 & 1 & x & 1 & x & x & x \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ x & x & 1 & x & x & x & x & x \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \end{matrix}\right\| \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} \end{array}$$

$$(f)\|P_{ki}\| = \left\|\begin{matrix} & & & 2 & & 2 & 2 & \\ & 2 & 2 & & & & & \\ & 2 & 2 & & 2 & & & \\ 2 & & & \cancel{4} & & 4 & 4 & \\ & & 2 & & & & & \\ 2 & & & 4 & & 4 & 4 & \\ 2 & & & 4 & & 4 & 4 & \\ & 4 & 4 & & 4 & & & \end{matrix}\right\| \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix}$$

In the result one term is disallowed as indicated by $\nmid$. The term may be readily identified as the minterm N_{33} from the octal indexes of row and column. By definition the minterm N_{33} is the short form of the input coincidence 011011≡33 in octal notation. Accordingly the function (f) may be regarded as the sum of two symmetric functions with the restriction of N_{33} on $\binom{6}{4}$

$$(f)=\binom{6}{2}+\binom{6}{4}N_{33}'$$

References

1. The Design of Switching Circuits (book), W. Keister, A. E. Ritchie, S. H. Washburn. D. Van Nostrand Company, Inc., Princeton, N. J., 1957, pp. 56–69.

2. Elements of Boolean Algebra for the Study of Information Handling Systems, R. Serrell. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, Oct. 1953, pp. 1366–80.

3. Matrix Analysis of Logical Networks, E. J. Schubert. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, Mar. 1958, pp. 10–13. (First of series)

4. Detection of Group Invariance or Total Symmetry of a Boolean Function, E. J. McCloskey, Jr. *Bell System Technical Journal*, New York, N. Y., vol. 35, no. 6, Nov. 1956, p. 1445.

5. Matrix Synthesis of High-Speed Logic, E. J. Schubert. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, Mar. 1959, pp. 4–8. (Second of series)

6. Analysis of Symmetric Logic, E. J. Schubert. *Report RM 58-76*, Burroughs Corporation, Research Center, Paoli, Pa. June 1958.

7. Matrix Algebra of Sequential Logic—Matrix Logic III, E. J. Schubert. *AIEE Transactions*, see pp. 1074–79 of this issue; also *Report TR 58-27*, Burroughs Corporation, section 6, July 1958.

8. Simultaneous Logical Equations—Matrix Logic IV, E. J. Schubert. *Ibid*, see pp. 1080–83 of this issue; also *Report TR 58-29*, Burroughs Corporation.

9. Magnetische Logische Grundschaltungen und Rechenanlagen, K. Ganzhorn. *Elektronische Rundschau*, Berlin-Borsigwalde, Germany, vol. 11, no. 8, Aug. 1957.

10. Transistor NOR Circuit Design, W. D. Rowe, G. H. Royer. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, July 1957, pp. 263–67.

Some Multihorned Dilemmas in the Magnetic Field

F. AVČIN
NONMEMBER AIEE

ENGINEERS do not like problems in units and dimensions if the discussion of them is likely to upset the concepts they learned in school. They usually shrug off such matters as "metaphysical." Unfortunately, there are enough theoretically minded scientists and engineers who insist on thinking and talking about such things to produce a continuing situation well represented by an old proverb: *Quot capita, tot sententiae* (From many heads, many opinions).

Nowhere is this better demonstrated than in the deliberations of Technical Committee 24, Electric and Magnetic Magnitudes and Units, International Electrotechnical Commission (IEC), particularly in relation to fundamental quantities that describe the magnetic field. The author has been closely involved in recent deliberations and has observed the epidemic of the malady *oersteditis acuta* which has spread among national committees.

It is the author's hope that the discussion which follows may clarify some of these problems, particularly with respect to those concerned with the magnetic field intensity H and its various cgs (centimeter-gram-second) and mksa (meter-kilogram-second-absolute) units, the oersted in particular. Perhaps there is here a lead to the cure of the disease. *Ne spes me faleret!* (Perhaps my hope might not deceive me!)

For this purpose it is necessary to adhere to a process of simple, careful exposition with numerical examples; complicated presentation, intricate logic, and especially lack of practical examples are to be avoided.

Background

Early experiments with magnetized needles led to Coulomb's equation defining the forces acting between point magnetic poles p:

$$\mathbf{F}=\mathrm{k}\frac{p_1p_2}{r^2}\mathbf{r}_0=p_1\mathbf{H} \tag{1}$$

where quantities are indicated in italics, measures in roman, and vectors in boldface. This equation also defines the magnetic field intensity H (also called, to the confusion of several generations of students, magnetic field strength and magnetizing force). This definition of $H=\mathrm{k}p/r^2$ is associated with the concept of "action at a distance," operating on a pole without the intermediary of the surrounding medium, like gravitation according to early concepts.

For magnetic fields produced by electric currents, equation 1 has its counterpart in the classical definition:

$$\mathbf{H}=-I\nabla\Omega \tag{2}$$

where Ω is the solid angle subtended, at the point where H is observed, by the current loop with current I. From equation 2 the value of H at the center of a long, uniformly wound solenoid can be directly derived:

$$H=4\pi\frac{NI}{l} \tag{3}$$

where N/l is the number of turns per unit length.

In 1882, Oliver Heaviside, a revolutionary in many areas of physics, mathematics, and engineering, realized that it is confusing (should we invent a word: "unrational"?) to find the factor 4π, which derives from spherical symmetry, in equation 3, which is concerned with a problem that might be called linear. He noted similar anomalies in other cases, both in electrostatic problems and electromagnetic problems. Therefore, he wrote equation 3:

$$H=\frac{NI}{l} \tag{4}$$

and, proceeding backward,

$$\mathbf{F}=\mathrm{k}\frac{p_1p_2}{4\pi r^2}\mathbf{r}_0 \tag{5}$$

This apparently arbitrary procedure leads to an interesting idea. From equation 5 it appears that

$$\mathbf{H}=\mathrm{k}\frac{p_2}{4\pi r^2}\mathbf{r}_0$$

represents the pole p_2 as having a (uni-

Paper 59-1187, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted December 3, 1958; made available for printing August 11, 1959.

F. Avčin is with the University of Ljubljana, Ljubljana, Yugoslavia.

The author is greatly indebted to Dr. Francis B. Silsbee, Chief, and Irwin L. Cooter, Raymond Sanford, and Chester Petersen, staff members, Electricity and Electronics Division, National Bureau of Standards, as well as to Dr. Carl C. Chambers, vice-president for Engineering Affairs, University of Pennsylvania, with whom he has discussed much of the Technical Committee 24, IEC, problems, and to Dr. S. Reid Warren, Jr., who gave this paper a digestible form.

This paper is derived from presentations at two colloquies held April 26, 1957, at the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa.; and April 29, 1957, at the National Bureau of Standards, Washington, D. C.

form) influence over a sphere of radius r; i.e., this is representative of the modern field concept involving the active participation of the medium, as contrasted with the obsolete action-at-a-distance concept.

Since the force $\mathbf{F}$ is the same, regardless of whether equation 1 or equation 5 is chosen, something has to give. Heaviside simply charged the units of magnetic pole strength by the factor $\sqrt{4\pi}$. In the cgs electromagnetic system, the unit pole in equation 1 is 1 cm (centimeter)$^{3/2}$-gm (gram)-second^{-1}, if k is dimensionless. Heaviside used the unit pole in equation 5 as $\sqrt{4\pi}$ cm$^{3/2}$-gm$^{1/2}$-second^{-1}, with k dimensionless. This was the first case of what is now called "rationalization by units."

Crux of the Problem

Now the trouble begins.

From equation 2 the classical definition of H gives $[H]=1$ abampere-steradian per centimeter (where [] is used to mean unit of), which, since there are 4π steradians in a unit sphere and 1 abampere equals 10 amperes, reduces to $10/4\pi$ ampere per cm. This nonrationalized unit of magnetic field intensity H is called the "oersted" by the majority of engineers and scientists, and so in most textbooks one finds:

$$1 \text{ oe (oersted)} = 10/4\pi \approx 0.8 \text{ ampere per cm (or ampere-turns per cm)}$$

This type of oersted may be called the "people's oersted" because of its popularity. It is dimensionally inconsistent with the fundamental units: centimeter-gram-second of the cgs system.

The dimensionally consistent and rationalized cgs unit of magnetic field intensity is 1 cm$^{-1/2}$-gm$^{1/2}$-second$^{-1}=10$ amperes per cm (or 10 ampere-turns per cm); it is 4π times larger than the people's oersted. It is believed that the IEC named this unit (10 amperes per cm) the "oersted" in 1930; however, eyewitnesses to that meeting differ in opinions except to agree that Oersted's name was bestowed upon the unit of the quantity "magnetic field intensity H," whatever it was or is.

Today we are led by modern concepts to look upon Coulomb's law of forces between magnetic poles as a useful approximation because we accept the hypothesis of equivalent Amperian currents rather than discrete magnetic poles. The reasons are clear: magnetic poles cannot be exactly located; they appear always in pairs; magnetic fields of elementary particles (including neutrons!) are always dipole fields; the forces on a magnet are spread throughout its volume; ferromagnetic polarization is primarily the result of electron spins, etc.

Thus it follows that H is, a priori, a theoretical mathematical tool. It is a figment of the imagination, in an important sense. It is true that whatever we do, we finally measure, not H, but

$$\mathbf{B}=\mu\mathbf{H}$$

the magnetic flux density (or magnetic induction) B, related to the magnetic field intensity H by the permeability μ. In the cgs 3-dimensional system of units, μ is a pure (dimensionless) number; in the mksa 4-dimensional system, μ is a physical quantity with measure and dimensions.

The torque $\mathbf{T}$ exerted on a magnetic dipole moment $\mathbf{m}$ can be written in two ways:

$$\mathbf{T}=\mathbf{m}_H\times\mathbf{H} \quad \text{or} \quad \mathbf{T}=\mathbf{m}_B\times\mathbf{B}$$

Now, if the magnetic moment $\mathbf{m}_H$ is defined in terms of the first relation, its unit is weber-meters; the unit of pole strength is webers, and this relation is, therefore, typical of the obsolete concept of point charges. On the other hand, the second relation gives $\mathbf{m}_B$ in ampere-meters^2, typical of the modern concept of Amperian currents (current loop).

Analysis of polarizable media likewise supports the concept that B is the realizable physical quantity, rather than H. In the rationalized system:

$$\mathbf{B}=\mu_0\mathbf{H}+\mathbf{J}$$

in which $\mathbf{J}$ is magnetic polarization (or intrinsic induction B_i). Whenever $\nabla\times\mathbf{J}$ exists there is an increase at a surface in the tangential component of $\mathbf{B}$. What is often called the demagnetization owing to an air gap in a magnetic circuit, is expressible, in this case, as the effect of the magnetic potential drop (ampere-turns) in the gap.

On the contrary, attempting to describe this situation in terms of H meets with logical difficulties. It then appears that poles on the ferromagnetic material, at the air gap, send $\mathbf{H}$ lines in one direction through the material and in the opposite direction through the air gap. Then the poles are attributed to $\nabla\cdot\mathbf{H}$, which produces an increase in the normal component of $\mathbf{H}$ in the air gap. The process of demagnetization is thus less easily understood.

The B field can be calculated from the distribution of uncanceled Amperian currents. The H field is computed from the corresponding distribution of pole strength. Mathematically, both concepts are equivalent; physically the B concept is the only reality, if anything of our descriptions of nature's behavior can be considered to be "real" for men's senses. But as the character of physical theories is essentially tentative and subject to experimental proof, only such quantities as can be measured have full claim on "reality."

All this may prove the intrinsic physical nonreality of the H. This is the first difficulty with, and the main objection to, the quantity "magnetic field intensity H."

The role of H is essentially that of an intermediary between the cause, the geometrical distribution of the exciting currents whether it is conductional or Amperian uncanceled, and the final effect, the physically real induction B. The excitation, the total current, alias the magnetic potential drop NI (ampere-turns) is, in some countries, such as Germany and Switzerland, denoted by the symbol Θ. Let the specific or linear excitation NI/l be symbolized by θ, for the purpose of this paper at least. With it we can write $H=\mathrm{k}\theta$, k being 4π in the nonrational and unity in the rational case. Using $B=\mu H$ we obtain:

$$B=\mu\mathrm{k}\theta$$

Equal causes, i.e., equal linear excitations θ, must naturally give equal magnetic effects and equal inductions B regardless of the definitional changes of the two intermediaries k and μ.

The logical step would be to try to live without such a trivial quantity as H, as the electrical engineer does. He plots his magnetization curves simply in the form $B=\mathrm{f}(\theta)$: gauss or kilolines per square inch versus the ampere-turns per inch or centimeter (see Proceeding *a*, Table I), according to an equation:

$$B=\mu\theta$$

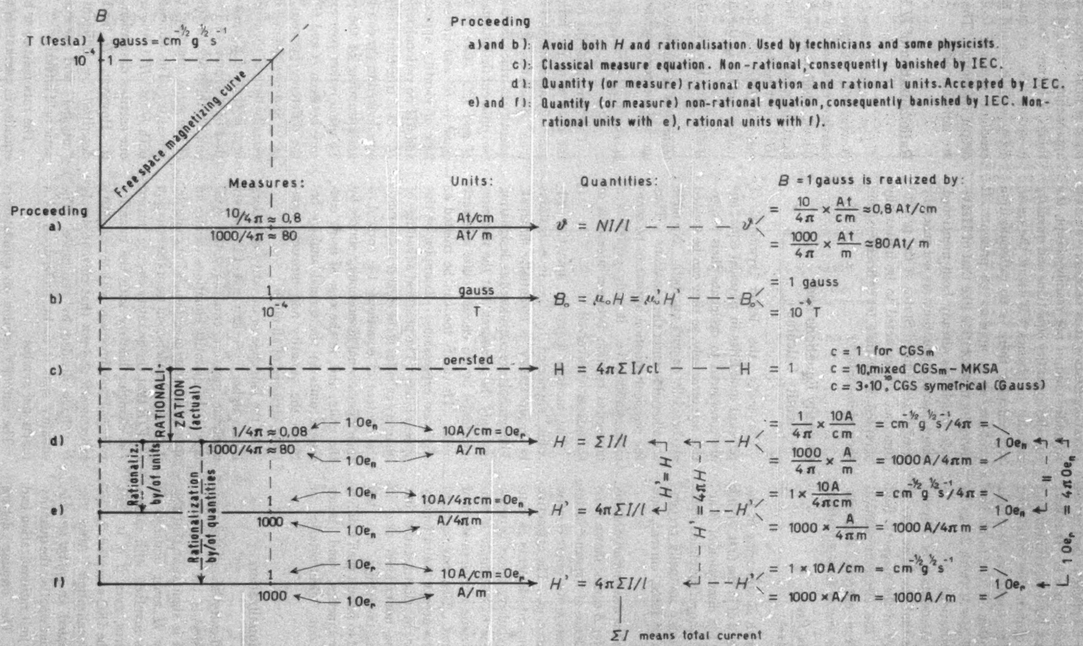
where θ is defined as being $\theta=NI/l$ for solenoid or toroid. More precisely it is the vector quantity:

$$\bar{\theta}=\oint\frac{I\,d\mathbf{s}\times\mathbf{r}_0}{4\pi r^2}$$

for more complicated current geometrics of the conduction and/or Amperian currents I. This $\bar{\theta}$ corresponds to the rationalized vector $\mathbf{H}$.

One argument in favor of H is its beautiful parallelism with the electric field strength $\mathbf{E}$ as defined from Coulomb's (Priestley's) law for electric point charges. But this parallelism is only a nice fiction, really only useful for freshmen in physics. The more elaborate theories of the electromagnetic field exhibit the relations $B-E$, $H-D$, instead of the parallelism $B-D$, $H-E$. This is particularly pertinent in

Table I. Realization of 1 Gauss in Standard Solenoid



Proceeding

a) and b): Avoid both H and rationalisation. Used by technicians and some physicists.

c): Classical measure equation. Non-rational, consequently banished by IEC.

d): Quantity (or measure) rational equation and rational units. Accepted by IEC.

e) and f): Quantity (or measure) non-rational equation, consequently banished by IEC. Non-rational units with e), rational units with f).

B
T (tesla)
gauss = cm$^{-1/2}$ g$^{1/2}$ s^{-1}
10^{-4}
1
Free space magnetizing curve
Proceeding
a)
b)
c)
d)
e)
f)
Measures:
$10/4\pi \approx 0{,}8$
$1000/4\pi \approx 80$
1
10^{-4}
$1/4\pi \approx 0{,}08$
$1000/4\pi \approx 80$
1
1000
1
1000
Units:
At/cm
At/m
gauss
T
oersted
10A/cm = Oe$_r$
A/m
10A/4πcm = Oe$_n$
A/4πm
10 A/cm = Oe$_r$
A/m
1 Oe$_n$
1 Oe$_n$
1 Oe$_n$
1 Oe$_n$
1 Oe$_r$
1 Oe$_r$
RATIONALIZATION (actual)
Rationaliz. by/of units
Rationalization by/of quantities
Quantities:
$\vartheta = NI/l$
$B_0 = \mu_0 H = \mu_0' H'$
$H = 4\pi \Sigma I/cl$
$H = \Sigma I/l$
$H' = 4\pi \Sigma I/l$
$H' = 4\pi \Sigma I/l$
$H' = H$
$H' = 4\pi H$
ΣI means total current
B = 1 gauss is realized by:
$\vartheta = \frac{10}{4\pi} \times \frac{At}{cm} \approx 0{,}8$ At/cm
$= \frac{1000}{4\pi} \times \frac{At}{m} \approx 80$ At/m
$B_0 = 1$ gauss
$= 10^{-4}$ T
$H = 1$
c = 1 for CGS$_m$
c = 10, mixed CGS$_m$ - MKSA
c = $3 \cdot 10^{10}$, CGS symetrical (Gauss)
$H = \frac{1}{4\pi} \times \frac{10A}{cm} = cm^{-1/2} g^{1/2} s^{-1}/4\pi =$
$= \frac{1000}{4\pi} \times \frac{A}{m} = 1000\,A/4\pi m =$
1 Oe$_n$
$H' = 1 \times \frac{10A}{4\pi cm} = cm^{-1/2} g^{1/2} s^{-1}/4\pi =$
$= 1000 \times \frac{A}{4\pi m} = 1000\,A/4\pi m =$
1 Oe$_n$
$H' = 1 \times 10\,A/cm = cm^{-1/2} g^{1/2} s^{-1} =$
$= 1000 \times A/m = 1000\,A/m =$
1 Oe$_r$
$= 4\pi$ Oe$_n$

Note: Procedure *d* corresponds to *d* and *c* in Table II; *e* corresponds to *e* and *f*; *f* corresponds to *a* and *b*. Also, with Procedure *e*, not 1 gauss but $1/4\pi$ gauss is realized (see Table II, Procedure *e*).

the relativistic Lorentz transformation and for the elegant Minkowski 4-dimensional vector presentation of the electromagnetic field, where, in Maxwell's equations, the pair $\mathbf{B}-\mathbf{E}$ appear as part of one vector equation and the pair $\mathbf{H}-\mathbf{D}$ as part of another. Thus both experiment and advanced theory confirm the correlation of B with E. The equations in media may thus be written in the form $D=\epsilon_0 E+p$ and $H=B/\mu_0-M$ (with $M=J/\mu_0$ being the magnetization).

The American Association of Physics Teachers suggests that the basic way to develop the theory of the magnetic field is to start with the electromechanical interaction between current elements. The corresponding vector force equation:

$$d\mathbf{F}=I d\mathbf{s}\times\mu_0 \oint \frac{I' d\mathbf{s}'\times \mathbf{r}_0}{4\pi r^2}$$

may be written with μ_0 appearing in the second factor instead of in the first, giving the second factor the dimension of B according to the well-known simplified equation $F=IlB$ for three mutually perpendicular directions of force, current, and field.

The factor 4π in the denominator of this rational equation for B can be shown to be the direct consequence of the same factor in Priestley's equation for point electric charges if put in motion. The attractive forces between current elements as contrasted with the repulsive forces between like charges appear to be simply a consequence of relativistic transformation of co-ordinate systems. This may serve to convince those who accept the rationally written equations in electrostatics but prefer to use the nonrational ones in magnetics. Both disciplines are too closely interlinked to permit such a differentiation. Electrokinetics, and hence magnetics, are but "electrostatics in motion."

Problem of Rationalization

The second difficulty about the magnetic field intensity is the rationalization problem. It is significant that this physical quantity H must undergo all the complications involved with rationalization procedures, even if it is "not real." In 1950 the IEC accepted the so-called total rationalization. After the voting, which created an *enfant terrible*, the question was put whether the quantity rationalization or unit rationalization had been accepted. The ensuing debate proved that there was no clear concept of what was meant by "total" rationalization and what type of rationalization is to be considered to be valid. So the 7-year-long rationalization war was started. The following paragraphs may help to settle it in its eighth year.

To understand clearly the booby traps of rationalization it is necessary at least to understand, if not readily accept, the so-called quantity equations and quantity calculus in physics. Without the quantity calculus there is no rationalization problem as it appears at present. It dwindles to a solution like that of Heaviside which has already been mentioned. Since quantity equations are not generally used in the United States at the present time but are expected to be universally accepted in due time, let us briefly explain their philosophy.

Quantity Equations and Quantity Calculus

Classically the letter symbols of physical quantities in physical equations are considered to designate their respective *measures* only, so they can be treated algebraically as pure numbers. No units are allowed to enter the equations. Thus the physical equation appear to be semi-mathematical, i e., it only expresses a numerical relation between measures of the quantities involved.

The modern approach, quantity calculus, supported by the investigations by group theory, is essentially different. The symbols are considered as designating not only the measures, but the measures multiplied by the chosen, quite arbitrary units. The units too can be treated as algebraic expressions. Examples of this appear later. The practical advantage of quantity calculus is the clearer insight into the content of equations; the control of their dimensional homogeneity simply by the units; and, in the case of using consistent units in the formal analogy between the quantity and measure equations, stripped of all units.

As the language of quantity equations is on a higher level than that of the purely algebraic measure equations, the symbols in the quantity equations require an appropriate interpretation. For example, it should be understood from the equation $I=V/R$ that the quantity electric current I is changing linearly with the quantity voltage V and inversely with the quantity resistance R and not that current simply *is* voltage divided by resistance. The equality sign is to be understood as a functional symbol.

The author's opinion on quantities is that 5 amperes and 10 amperes are the *same quantity:* the electric current, but of two different magnitudes.

One must not forget that men use physical quantities and then try to depict by images the behavior of nature, while their ultimate reality remains ever unknown. Each successive page in nature's great book is more difficult to turn. We are approaching the ultimate essence of things, it is true, but asymptotically only it seems. So, as quantity calculus is doing well, it should be accepted as an excellent working tool, without much deeper reasoning. The depths of nature are unfathomable.

Rationalization of H by Units

The reasoning involved is much like Heaviside's. If, from the nonrational measure equation, a measure H' is obtained, which is 4π times larger than that from the rational H, the nonrational unit $[H']$ must be 4π times smaller than the rational $[H]$. The quantities H and H' thus are said to be the same:

$$H=\mathrm{H}\cdot[H]=\mathrm{H}'\cdot[H']=H'$$

regardless of the way of writing the numerical measure equation of H:

$$\mathrm{H}=\mathrm{NI}/l \quad \text{or} \quad \mathrm{H}'=4\pi\mathrm{NI}/l$$

This principle, the ratio of units being the inverse of the ratio of measures, is generally accepted as if it were universally true. However, it is not valid if the quantity to which the two units belong has two different definitions as is the case with H and H'. Here H is not identically equal to H' and the consequence of the lack of identity appears in Tables I, II, and III on rationalization procedures for H. It is true that this point is generally over-looked to the detriment of the proper understanding of the problem involved. To use a rather poor example, the difference between H and H' is comparable to the difference between a surface defined by r^2 and one defined by $4\pi r^2$. Dimensionally both of them are surfaces but the first is that of a square and the second that of a sphere 4π times larger. In H and H' there are two different definitions of the same quantity: the magnetic field intensity. The dimensional equation of both H and H' is the same $[\mathrm{IL}^{-1}]$ unless one ascribes dimension to the solid angle Ω. However, H and H' should be described by different adjectives: the "rational" and the "nonrational" magnetic field intensity.

The rationalization by units results in equality in amounts of the two differently defined magnetic field intensities. It can be seen from Table II that despite the fact that both H and $H'=1$ amp/m (ampere per meter), the corresponding

Vol.
Pt. 1
1959

Table II. Rationalization Procedures for Magnetic Field Intensity H in Light of Quantity Equations

1	2	3	4	5	6	7	8	9	10	11	12	Remarks
Procedure	Rationalization of equation defining H	Measure system	Quantity equation defining the quantity "magnetic field intensity" $H = \mathrm{H}\cdot[H]$	Measure H of the quantity H H	Unit $[H]$ of the quantity H $[H]$	Unit $[I]$ of current I $[I]$	Unit $[l]$ of lenght $[l]$	Realization of standard: Linear excitation ϑ to apply to solenoid in order to realize $H = 1\cdot[H]$ $\vartheta = \frac{N}{l}I = \frac{\Theta}{l}$	Permeability μ_0 of free space necessary to obtain $B_0 = \mu_0 H$ $\mu_0 = \mu_0[\mu_0]$	Induction B_0 in solenoid at realizing $H = 1\cdot[H]$ with ϑ in column 9 B_0	Permittivity ϵ_0 of free space $\epsilon_0 = 1/\mu_0 c_0^2$	Quantities are in *Italics*, Measures and numbers are in ROMAN type.
a	by units (down) $\mathrm{H}\cdot[H] \equiv \mathrm{H}'\cdot[H']$	CGS_m	USUAL NON-RATIONAL PROCEDURE BUT IN MEASURE-FORM ONLY $H' = 4\pi\overbrace{\frac{N}{l}I}^{\vartheta} = 4\pi\frac{NI}{l}\times$ non-rational		$\times\frac{10A}{4\pi cm} \triangleq 1\frac{cm^{-1/2}g^{1/2}s^{-1}}{4\pi}$ "people's" oersted (n.r.) Oe_n	$1\,cm^{1/2}g^{1/2}s^{-1} \triangleq$ $\triangleq 10\,A$	1 cm	$\frac{1}{(4\pi)^2}\cdot\frac{10\,At}{cm} = \frac{1}{4\pi}\cdot\frac{10\,At}{4\pi cm}$	$\mu_0' = 1 \equiv$ $\equiv \frac{1}{4\pi}\times\frac{G}{Oe_n}$	$\boxed{B_0 = \frac{1}{4\pi}G \text{ for } H' = 1\cdot Oe_n}$	$\epsilon_0' = \frac{10^{-20}}{9}\times cm^{-2}s^2 \equiv$ $\equiv 4\pi\cdot\frac{10^{-20}}{9}\times cm^{-2}s^2\cdot\frac{Oe_n}{G}$	The non-rational and non-consistent unit $10A/4\pi cm \triangleq \frac{1}{4\pi}cm^{-1/2}g^{1/2}s^{-1}$ is universally considered to be "the oersted" ("people's oersted") Oe_n: $1\cdot Oe_n \triangleq 80\,A/m$
b		MKSA	magnitudes equal $\frac{NI}{l}\cdot\frac{A}{m} = \frac{4\pi NI}{l}\cdot\frac{A}{4\pi m}$ $\frac{NI}{l}\cdot\frac{10A}{cm} = \frac{4\pi NI}{l}\cdot\frac{10A}{4\pi cm}$ $H \triangleq H'$		$\times\frac{A}{4\pi m} \triangleq 10^{-3}\,Oe_n$ MKSA - oersted (n.r.)	1A	1m	$\frac{1}{(4\pi)^2}\cdot\frac{At}{1m}$	$\mu_0' = 10^{-7}\times\frac{T}{Am^{-1}} \equiv$ $\equiv \frac{10^{-7}}{4\pi}\times\frac{T}{A/4\pi m}$	$B_0 = \frac{10^{-7}}{4\pi}\cdot T$ for $H' = 1\cdot A/4\pi m$	$\epsilon_0' = \frac{10^{-9}}{9}\times\frac{As}{Vm}$	The non-rational and non-consistent unit $A/4\pi m$ is sometimes named "the milli-oersted" or "MKSA-oersted". It is seldom used. 1 gauss (G) = $1\,cm^{-1/2}g^{1/2}s^{-1}$
c		MKSA	USUAL RATIONAL PROCEDURE $H = \overbrace{\frac{N}{l}I}^{\vartheta} = \frac{NI}{l}\times$ rational		$\times\frac{A}{m} \triangleq 10^{-3}\,Oe_r$ MKSA (Giorgi) - unit (r.)	1A	1m	$1\frac{At}{m} = 10^{-3}\cdot\frac{10\,At}{cm}$ $(4\pi)^2 : 1$	$\mu_0 = 4\pi 10^{-7}\times\frac{T}{Am^{-1}}$	$(4\pi)^2 : 1$ $\boxed{B_0 = 4\pi 10^{-7}\cdot T \text{ for } H = 1\cdot A/m}$	$\epsilon_0 = \frac{10^{-9}}{36\pi}\times\frac{As}{Vm}$	1 T (tesla) $\triangleq 10^4$ gauss is the new IEC unit of B. $T = Nm^{-1}A^{-1} = Vs/m^2 = Wb/m^2 \triangleq 10^4\,G$ With "turn" in At and $\frac{N}{l}$ no angle-concept is involved.
d	Rationalization of quantity H by measures $\mathrm{H}''[H''] \neq \mathrm{H}[H]$	CGS_m	magnitudes non-equal $\frac{4\pi NI}{l}\cdot\frac{A}{m} \neq \frac{NI}{l}\cdot\frac{A}{m}$ $\frac{4\pi NI}{l}\cdot\frac{10A}{cm} \neq \frac{NI}{l}\cdot\frac{10A}{cm}$ $H'' \neq H$		$\times\frac{10A}{cm} = 1\,cm^{-1/2}g^{1/2}s^{-1}$ IEC - oersted (r.) Oe_r	$1\,cm^{1/2}g^{1/2}s^{-1} \triangleq$ $\triangleq 10\,A$	1 cm	$\frac{10\,At}{cm} = 10^3\frac{At}{m}$	$\mu_0 = 4\pi \equiv$ $\equiv 4\pi\times\frac{G}{Oe_r}$	$B_0 = 4\pi\cdot G$ for $H = 1\cdot Oe_r$	$\epsilon_0 = \frac{10^{-20}}{36\pi}\times cm^{-2}s^2$	The rational and consistent unit $1\,cm^{-1/2}g^{1/2}s^{-1} \triangleq 10\,A/cm$ is the (rational) oersted Oe_r as defined within the IEC. It is seldom used. $1\cdot Oe_r \triangleq 1000\,A/m$
e		CGS_m	$H'' = 4\pi\overbrace{\frac{N}{l}I}^{\vartheta} = 4\pi\frac{NI}{l}\times$ non-rational	$[H''] = [H]$	$\times\frac{10A}{cm} = Oe_r$	$1\,cm^{1/2}g^{1/2}s^{-1} \triangleq$ $\triangleq 10\,A$	1cm	$\frac{1}{4\pi}\cdot\frac{10\,At}{cm} \approx 80\frac{At}{m}$ $1/4\pi : 1$	$\mu_0'' = 1 \equiv$ $\equiv 1\times\frac{G}{Oe_r}$	$1/4\pi : 1$ $\neq$ $=$ $\boxed{B_0 = 1\cdot G \text{ for } H'' = 1\cdot Oe_r}$	$\epsilon_0'' = \frac{10^{-20}}{9}\times cm^{-2}s^2 \equiv$ $\equiv \frac{10^{-20}}{9}\times cm^{-2}s^2\cdot\frac{Oe_n}{G}$	The quantity B is never rationalized. So the values of μ_0 and ϵ_0 are determined from $B_0 = \mu_0 H_0 = \mu_0' H' = \mu_0'' H'' = \dots$ and $\epsilon_0\mu_0 c_0^2 = \epsilon_0'\mu_0' c_0^2 = \epsilon_0''\mu_0'' c_0^2 = 1$
f	by units (up) $\mathrm{H}'''\cdot[H'''] = \mathrm{H}''\cdot[H'']$	MKSA	magnitudes equal $\frac{NI}{l}\cdot\frac{4\pi A}{m} = \frac{4\pi NI}{l}\cdot\frac{A}{m}$ $\frac{NI}{l}\cdot\frac{4\pi 10A}{cm} = \frac{4\pi NI}{l}\cdot\frac{10A}{cm}$ $H''' \triangleq H''$		$\times\frac{A}{m} = 10\,Oe_r$	1A	1m	$\frac{1}{4\pi}\cdot\frac{At}{m}$	$\mu_0'' = 10^{-7}\times\frac{T}{Am^{-1}}$	$B_0 = 10^{-7}\cdot T$ for $H'' = 1\cdot A/m$	$\epsilon_0'' = \frac{10^{-9}}{9}\times\frac{As}{Vm}$	The consistent unit of permeability is $T/Am^{-1} = Vs/Am = H/m = N/A^2$. If practical MKSA-units be identified with decadic multiples of CGS_m-units one obtains: $1\,T/Am^{-1} \equiv 10^7$
g		MKSA	$H''' = \overbrace{\frac{N}{l}I}^{\vartheta} = \frac{NI}{l}\times$ rational		$\times 4\pi\frac{A}{m} \triangleq 10\,Oe_n'''$	1A	1m	$4\pi\frac{At}{m}$ $(4\pi)^2 : 1$	$\mu_0''' = 4\pi 10^{-7}\times\frac{T}{Am^{-1}} \equiv$ $\equiv (4\pi)^2 10^{-7}\times\frac{T}{4\pi Am^{-1}}$	$(4\pi)^2 : 1$ $B_0 = (4\pi)^2 10^{-7}\cdot T$ for $H''' = 1\cdot 4\pi\,A/m$	$\epsilon_0''' = \frac{10^{-9}}{36\pi}\times\frac{As}{Vm}$	This type of rationalization by units (up) has never been used. The non-rational and non-consistent unit $4\pi\,cm^{-1/2}g^{1/2}s^{-1}$ would represent its "oersted" $Oe_n''' = 4\pi Oe_r = (4\pi)^2 Oe_n$ Likewise for the corresponding MKSA-unit $4\pi Am^{-1}$
h		CGS_m			$\times 4\pi\frac{10A}{cm} = 1\cdot Oe_n''' \equiv$ $(\equiv 4\pi\,cm^{-1/2}g^{1/2}s^{-1})$	$1\,cm^{1/2}g^{1/2}s^{-1} \triangleq$ $\triangleq 10\,A$	1cm	$4\pi\cdot\frac{10\,At}{cm}$	$\mu_0''' = 4\pi \equiv$ $\equiv (4\pi)^2\times\frac{G}{Oe_n'''}$	$B_0 = (4\pi)^2\cdot G$ for $H''' = 1\cdot Oe_n'''$	$\epsilon_0''' = \frac{10^{-20}}{36\pi}\times cm^{-2}s^2$	

The Gaussian $H = \frac{4\pi}{c}\frac{NI}{l}$ and the "technical" $H = \frac{4\pi}{10}\frac{NI}{l}$ are not quantity - but measure - equations only.

≙ read "corresponds to"
≠ read "non-equal to"
≡ read "identically equal to"

Table III

Linear Excitation	At/m	At/Cm	At/Inch	Oersted
1 At/m =	1	10^{-2}	0.0254	$4\pi/1{,}000 \approx 0.0126$
1 At/cm =	10^2	1	2.54	$4\pi/10 \approx 1.26$
1 At/inch =	$39 \approx 40$	$0.39 \approx 0.4$	1	$0.49 \approx 0.5$
1 oe $\triangleq$	$1{,}000/4\pi \approx 80$	$10/4\pi \approx 0.8$	$2.02 \approx 2$	1

flux densities B_0 are not equal, being actually $4\pi \cdot 10^{-7}$T (tesla*) and 10^{-7}T, as if there were two meanings of the unit amp/m possible (see Procedures *b* and *c* Table II). In fact, the unit "ampere per meter," cannot claim much illustriousness. Thus units rationalization cannot claim much reality in the quantity sense. Equal $H=H'$ ought to give equal B_0, but it does not.

If the rational units of H are 10 amp/cm (cgs$_m$, IEC oersted) and 1 amp/m (mksa), the nonrational and consequently nonconsistent, ones are 10 amp/4πcm (people's oersted) and 1 amp/4πm (millipeople's oersted, mksa oersted).

This type of rationalization is called rationalization of/or by units, though what is primarily rationalized is always the equations relating certain quantities.

Rationalization by Quantities

People who see true quantities in the symbols NI/l and H of the definitional equation for H, and not merely their measures in a chosen system of units, and who also use consistent units, are not in favor of sacrificing the double advantages of quantity equations used with consistent units. They prefer to leave the units of H and H' unchanged and, instead, describe the same natural phenomenon, the same magnetic field by two related-in-definition magnetic field intensities: the rational $H=NI/l$ and the nonrational $H'=4\pi NI/l$, saying that $H'=4\pi H$. (Note that in Table II the H″ corresponds to H′ here.)

With the unit 1 amp/m for both H and H', they would have, for example, $H=$ 5 amp/m and $H'=20\pi$amp/m with the *same* B_0 in both cases (Procedures *c–f* in Table II).

$$B_0=4\pi 10^{-7}\frac{\mathrm{T}}{\mathrm{Am}^{-1}}\times 5\mathrm{Am}^{-1}=10^{-7}\frac{\mathrm{T}}{\mathrm{Am}^{-1}}\times 20\pi\mathrm{Am}^{-1}=2\pi 10^{-6}\mathrm{T}$$

where A = amp. Thus the quantity, absolute permeability μ_0, must conform to the rational or nonrational H used. With $\mu_0'=(1/4\pi)\mu_0$ and $H'=4\pi H$:

$$B_0=\mu_0 H=\mu_0' H'$$

The intermediary μ_0 between the invariant reality B_0 and the auxiliary nonreal H must adapt itself to the changing definition of H.

Papers written on the subject of the comparison of rationalization by units and quantities usually discuss two equal versus two different quantities H. We have seen that in both cases the measures do change. In the first case the unit is also changed, while in the second it is unchanged. But H and $H'=4\pi H$ give equal B_0, as required, so the rationalization by "quantities" can claim physical "reality" to the extent that H can.

One of the consequences is a confusion in the equations containing the magnetic moment, a quantity particularly important in atomic and nuclear physics. We have seen that there can be a B moment and an H moment. As the latter can be combined with both the rational and nonrational magnetic field intensity the ensuing confusion in atomic constants is quite appreciable, e.g., various values and dimensions of Bohr's magneton, gyromagnetic ratios, etc. But the practical engineer always uses the B moment while calculating his electric machines. He must be quite sure of himself as his machinery must work exactly as rated.

If we, like he, exclude, in magnetics at least, the intermediary H, the linear excitation θ of conduction and/or Amperian currents and their effects, the induction in media $B=B_0+J$ (B_0 in air plus the contribution of the medium, the magnetic polarization J or the intrinsic induction B_i), all the enumerated difficulties immediately cease to exist. The free space permeability μ_0 can have but one unit and value, the rational one. However, there will always be cases when H can be very illustrative, e.g., in electromagnetic waves in the Poynting vector of power density flow:

$$\mathbf{P}=\mathbf{E}\times\mathbf{H}=\mathbf{E}\times\mathbf{H}'/4\pi$$

though this can always be written in the invariant B form:

$$\mathbf{P}=\mathbf{E}\times\mathbf{B}/\mu_0$$

With the newly accepted IEC unit of B, the tesla, T (1 T = 1 vs (volt-second)/m^2 = 1 weber/m^2 = 10^4 g [gauss]) of the mksa system, and the old cgs$_m$ unit (1 g = 1 cm$^{-1/2}$-gm$^{1/2}$-second^{-1}) and the corresponding units of θ being 1 amp-t (ampere-turn)/m and 10 amp-t/cm, the consistent unit of μ_0 is always $[\mu_0]=[B_0]/[\theta]$:

$$[\mu_0]_{\mathrm{mksa}}=\mathrm{T/Atm^{-1}}$$

and

$$[\mu_0]_{\mathrm{cgs_m}}=\mathrm{g/10\ Atcm^{-1}}$$

In the latter unit 10 amp-t/cm with 10 amps = 1 abampere (not "absolute ampere" but the current unit 1cm$^{1/2}$-gm$^{1/2}$-second^{-1} of the absolute electromagnetic cgs system) corresponds in cgs dimension to the gauss, so that $[\mu_0]=1$. If H is not omitted, the mksa and cgs$_m$ units of permeability are as presented in column *10* of Table II.

The numerical value of μ_0 is readily obtained from the current balance equation which is at the base of the definition of the absolute ampere:

$$F=\mu_0 lI^2\mathrm{t}/2\pi r$$

with $l=1$ m, $r=1$ m, $I=1$ amp, one turn (t) and $F=2\cdot 10^{-7}$, newton = 0 02 dyne (newton N = kgms^{-2} [kilogram-meter/(second)^2] = joule/m one has, as an example of the quantity calculus:

$$\mu_0=\frac{2\pi Fr}{lI^2\mathrm{t}}=\frac{2\pi\times 2\cdot 10^{-7}\mathrm{N}\times 1\mathrm{m}}{1\mathrm{m}\times 1\mathrm{A}^2\times 1\ \mathrm{turn}}=4\pi\cdot 10^{-7}\frac{\mathrm{newton}}{\mathrm{amp^2\text{-}t}}$$

or with

$$\frac{\mathrm{N}}{\mathrm{A}^2}=\frac{\mathrm{VAsm}^{-1}}{\mathrm{A}^2}\times\frac{\mathrm{m}^2}{\mathrm{m}^2}=\mathrm{Vsm^{-2}/Am^{-1}}=\mathrm{T/Am^{-1}}$$

and finally,

$$\mu_0=4\pi\cdot 10^{-7}\ \mathrm{T/Atm^{-1}}$$

corresponding to the units T of B and amp-t/m of θ in the equation $B=\mu_0\theta$. The corresponding, always rational, value of μ_0 in the cgs system is quickly obtained if the 4-dimensional mksa Giorgi system is allowed to degenerate to the merely 3-dimensional cgs$_m$ electromagnetic system. This immediately happens if the mksa units are identified with the usual decimal multiples of the corresponding cgs$_m$ units. If we put 1 amp $\equiv 10^{-1}$ cm$^{1/2}$-gm$^{1/2}$-second^{-1} and 1T = 10^{-4}-cm$^{1/2}$-gm$^{1/2}$-second^{-1} we have, with 1 m = 10^2 cm:

$$\frac{\mathrm{T}}{\mathrm{Atm}^{-1}}=\frac{10^4\ \mathrm{cm^{-1/2}\text{-}gm^{1/2}\text{-}second^{-1}}}{10^{-1}\ \mathrm{cm^{1/2}\text{-}gm^{1/2}\text{-}second^{-1}}\times(10^2\ \mathrm{cm})^{-1}}\equiv 10^7$$

and $\mu_0=4\pi$, a pure number.

---

* The IEC is making a valiant effort to establish the tesla (abbreviation T) as the mksa unit of magnetic flux density; 1 tesla is 1 weber per square meter.

The absolute ampere, the fourth basic unit of the mksa system, and all units derived from it may consequently be considered to be defined independently of the cgs system. Alternatively, if not so, one has:

$$1 \text{ amp} \equiv 10^{-7/2} \text{ kg}^{1/2}\text{-m}^{1/2}\text{-second}^{-1}$$

and the ampere becomes a unit derived from the three basic units of the mks system of mechanics. Four-dimensional systems are simpler than 3-dimensional systems, from both the mathematical and the formalistic-aesthetic points of view. Hence the equation: 1 ampere is 0.1 of the cgs_m abampere, is to be read as a correspondence only instead of an equality:

$$1 \text{ amp} \hat{=} 10^{-1} \text{ cm}^{1/2}\text{-gm}^{1/2}\text{-second}^{-1}$$

Thus the induction 1 tesla and 1 gauss, respectively, are obtained in air by a local linear excitation $\theta = B/\mu_0$ equal to, for tesla:

$$\theta = \frac{1\text{T}}{4\pi 10^{-7}\text{T/Atm}^{-1}} = \frac{10^7}{4\pi} \cdot \frac{\text{At}}{\text{m}}$$

and for gauss:

$$\theta = \frac{1\text{g}}{4\pi} = \frac{1}{4\pi} \cdot \frac{10\text{At}}{\text{cm}} \approx 0.8 \frac{\text{At}}{\text{cm}}$$

The important result is that it is 1 g which is obtained by the linear excitation 0.8 amp-t/cm, a statement so similar to the ancient equality between the classical people's oersted and 0.8 amp/cm. Tables I and II illustrate this still more clearly, with the standard solenoid used to realize the variable units of H and the stable units of B.

Realization of Units of *H* and *B* in Standard Solenoid

Quantities are our images of ultimate realities in physics. As every physical theory is of little use unless proved by experiment, quantities without a physical reality, i.e., not accessible to measurement, would be of little use. For every quantity there must be a standard permitting the quantity to be realized with the desired amount, which is often exactly one chosen unit of the quantity. Therefore, standards are material devices which must be somehow energized to produce a quantity of a certain amount for the purpose of measurements, i.e., of comparing. The international meter or, better still, the meter together with a precise length comparator, is such a material primary standard as are the current balance, the apparatus of the time service of any astronomical observatory, etc. Their precision accounts for the ultimate accuracy of measurements obtainable by such primary standards.

The standard for the magnetic field is an exactly known geometrical arrangement of wire with an exactly known current in it. "Exactly" in standards means that the tolerance in geometry and current is estimated as exactly as required, absolute certainty being the privilege of uneducated minds. The standard of the people's oersted has been suggested as the circular arc of wire of 1-cm arc length at a radius of 1 cm with two radial leads along the axis. The axial magnetic field strength at the center of the arc is then 1 oersted. If the whole circle is taken, it is 2π oersteds. This standard of the nonrational oersted has the disadvantage of being quite unrational indeed. It gives a low field and a very nonuniform one. The only rational standard, encountered in every well-equipped laboratory, is the precision-built standard solenoid of well-defined pitch l/N (length for a turn). Energized by the current I it produces the linear excitation $\theta = IN/l$ If it is sufficiently long, and uniformly and densely wound, the linear excitation and, so, the field in it, is highly homogeneous over a large volume and can be calculated. Therefore further discussion of the realization of units of H and B is based on the standard solenoid.

As stated already, the same θ will always produce the same B. H, on the contrary, will appear according to its chosen definition equation and unit; μ_0 will have to adapt itself to H in order to give the invariant B. Table I will show, if carefully studied, what is realized, and how, including the trickery of rationalization procedures on eight examples: four mksa and four cgs_m, all based on the physically advanced concept of quantity equations. The table shows:

1. The rationalization of the quantity magnetic field intensity by units: transition from the nonrational H' in rows *a–b* to the rational H in rows *c–d*, so that there is equality in magnitude $\text{H}'[H'] = \text{H}[H]$; transition from nonrational H'' in rows *e–f* to rational H''' in lines *g–h*, giving similarly $\text{H}'''[H'''] = \text{H}''[H'']$. The latter procedure, *g–h*, is never encountered, fortunately, but it is equally possible and analogous to the first. It is accordingly added as illustration. The units involved are in column *6*.

2. The rationalization of the quantity magnetic field intensity by measures (which is usually said to be rationalization of quantities): transition from the nonrational H'' in rows *e–f* to the rational H in rows *c–d*. There is no longer equality in magnitude as $\text{H}''[H''] \neq \text{H}[H]$. The two magnetic field strengths H'' and H, different in definition, do not have equal magnitudes as is the case with the two procedures for rationalization by units under item 1.

3. If the unit $[H]$ in column *6* of the quantity H as defined in column *4*, is to be "realized," the linear excitation $\theta = NI/l$ as in column *9* must be applied to an appropriate solenoid. For proceeding in row *a*, for example, we must realize the people's oersted. This is obtained if $H' = 1 \times 10$ amp/4πcm. If, in the corresponding definitional equation $H' = 4\pi NI/l$, we substitute for

$$NI/l = \theta = \frac{1}{(4\pi)^2} \cdot \frac{10 \text{ At}}{\text{cm}} ,$$

and have, using quantity procedure,

$$H' = 4\pi \cdot NI/l = 4\pi \cdot \frac{1}{(4\pi)^2} \cdot \frac{10\text{At}}{\text{cm}} = 1 \cdot \frac{10 \text{ A}}{4\pi\text{cm}}$$

as required. The same procedure is evident for all the eight cases.

We see that θ changes in the ratio $(4\pi)^2$ to 1 for the two cases of rationalization by units and in the ratio 4π to 1 in the case of rationalization by measures (quantities).

It is also noted that only with rationally written equations do the linear excitation θ and the unit $[H]$ correspond, except for the factor turn (t).

4. The induction B_0 in the solenoid, corresponding to the θ in column *9*, is evident from column *11*. Naturally B_0 goes in parallel with θ, changing in steps $(4\pi)^2$ to 1 for the two rationalizations by units and in steps 4π to 1 in the rationalization by measures (quantities).

5. The corresponding μ_0 which relates the "realized" unit-magnetic field intensities $H = 1 \cdot [H]$ to the actually physically existing inductions $B_0 = \mu_0 H$, are evident from column *10*. Their units $[\mu_0]$ are also given as the ratio $[B]$ to $[H]$. In the four cgs_m cases the pure number obtained for μ_0 is further illustrated with the unit gauss/oersted. This may imply the physically different character of B and H, even in cgs systems and free space. If B and H are recognized to differ essentially in matter, they are logically different in free space too, despite the fact that $\mu_0 = 1$ in Procedure *e*, Table II, as the transition from matter into free space can be considered continuous. Also they are always plotted on two different axes of the co-ordinate system.

6. Intercomparison of realized B while $H = 1 \cdot [H]$, both evident from column *11*, is startling. However, it clearly discloses some inherent difficulties in understanding H and its rationalization procedures.

In both rows *d* and *e* the IEC oersted $1 \cdot \text{oe}_r = 10$ amp/cm is realized. But as θ is different, B must be too. It is 4πg (gauss) and $1 \cdot$g respectively; the same is true for the $1 \cdot$amp/m in rows *c* and *f*, which gives $4\pi \cdot 10^{-7}$T and 10^{-7}T.

The proper understanding is simple. The oe_r and the amp/m are units of the two differently defined magnetic field strengths H and H'', both using the same unit, equal in magnitude but not identical in character. This is no more difficult to understand than the attempt to describe some physically real B by two amounts of magnetic field strength differing in magnitude by a factor 4π. The literature thus speaks of two "paradoxical" meanings of amp/m:

$$\left(\frac{\mathrm{A}}{\mathrm{m}}\right)^{\mathrm{i}}=\frac{1}{4\pi}\left(\frac{\mathrm{A}}{\mathrm{m}}\right)^{\mathrm{r}}$$

and comes to such conclusions, on a higher level of abstraction, as:

$$\left(\frac{\mathrm{cm}}{\mathrm{cm}}\right)^{\mathrm{i}}=\frac{1}{4\pi}\left(\frac{\mathrm{cm}}{\mathrm{cm}}\right)^{\mathrm{r}}$$

or vice versa. They no doubt require a rather high level of imagination and are indigestible to the realist in electrical engineering.

Based on the philosophy included in Table II, we can finally answer the two IEC questions: 1. What type of rationalization has been or is to be chosen? 2. What is the oersted, and how much field does it represent? All that is still needed is an insight into how the old-timers have been proceeding while calculating their magnetic fields. Their procedure may be explained by the example of the superhigh H parameter the author is designing for the Magnetic Measurements Section of the National Bureau of Standards: "20,000 oe at least are required. The gap is to be 3 cm long and 5 cm in diameter."

Old-timers have their result immediately: 20,000 oe require 16,000 amp-t/cm as 1 oe = 0.8 amp-t/cm. He who does not believe may look it up in any textbook or handbook. If you ask for the induction B, you get the answer: as $\mu=1$ in air, B is 20,000 g, naturally, 1 g being 1 oe. This is all.

If they were asked what field they obtained from 16,000 amp-t/cm, they use the measure equation $H=4\pi NI/10l$.

With $NI/l=16{,}000$ it gives 20,000. What? Oersteds naturally.

Very simple and correct, correct in its final result at least, as their machinery and apparatus prove. From the logic of quantity calculus however, there are two main flaws. The first is that by saying 1 oe = 10 amp-t/4πcm the units of two different quantities, H and θ, are being simply identified. The second is similar. By writing 1 oe = 1 g, a unit of H is identified with a unit of B. But this equation is never meant so seriously. In reality only measures are being compared: The number of gausses is equal to the number of oersteds. The classical procedure is essentially numerical and nonquantity. The so-called "technical" (nonrational) equation $H=4\pi NI/l01$ and the Gaussian (equally nonrational) $H=4\pi NI/\mathrm{c}l$, so beloved of some physicists, are typical measure equations, as the factors 10 and $\mathrm{c}=3\cdot 10^{10}$ (the measure of the velocity of light in the unit cm/second) came into them by a particular choice of units for the quantities involved, while true quantity equations are usable for all possible units if they too are treated algebraically.

In the Gaussian equations c appears in such quasi-static cases as above and disappears from equations where it should appear and represent the velocity of light c_0, e.g., for wavelength of an oscillator:

$$\text{Gaussian } \lambda=2\pi\sqrt{\mathrm{LC}}$$

$$\text{Quantity } \lambda=2\pi c_0\sqrt{LC}$$

It is absurd that in the Gaussian symmetric system all three quantities; wavelength, inductance, and capacitance, are to be measured in the same unit, the centimeter.

Let us quickly see how permeability is calculated, namely, the relative permeability μ_r, the only important one as it appears in tables, magnetization curves, etc. It tells how many times the magnetic field in matter is larger than in free space or air, equal linear excitation assumed. In the Armco Steel Corporation oriented-grain *M-6* silicon steel, $\theta=1$ amp-t/cm gives $B=16{,}000$ g.

Classical measure procedure: with 1 amp-t/cm = 1.26 oe ($1.26\approx 4\pi/10$) the relative permeability is:

$$\mu_r=\frac{16{,}000}{1.26}=12{,}800$$

Cgs quantity procedure:

$$\mu_r=\frac{B}{H}=\frac{16{,}000\ \mathrm{g}}{1.26\ \mathrm{oe}}$$

Using quantity calculus this is the pure number 12,800 only if 1 g = 1 oe, i.e., the 1 oe is the IEC oersted, so that:

$$\mu_r=\frac{16{,}000\ \mathrm{g}}{1.26\ \mathrm{g}}=12{,}600$$

Mksa quantity procedure: $B=16{,}000\ \mathrm{g}=1.6\ \mathrm{T}$ and $\theta=1$ amp-t/cm gives $B_0=1.26\ \mathrm{g}=1.26\cdot 10^{-4}\ \mathrm{T}$. So:

$$\mu_r=\frac{B}{B_0}=\frac{1.6\ \mathrm{T}}{1.26\cdot 10^{-4}\ \mathrm{T}}=12{,}800$$

and the absolute permeability:

$$\mu_0=\frac{B_0}{\theta}=\frac{4\pi\cdot 10^{-5}\ \mathrm{T}}{10^2\ \mathrm{Atm}^{-1}}=4\pi\cdot 10^{-7}\ \mathrm{T/Atm}^{-1}$$

as seen from row *c*, column *10* of Table II.

As these two examples of classical and illogical procedure have no effect on the final physical result, nobody cared much about them until 1950. But they may help to answer the two IEC questions:

1. What type of rationalization, quantity or unit, is to be chosen?

The "actual" rationalization is the transition of classical nonrational "measure" definitional equation for H, combined with nonrational nonconsistent units (Procedures *a–b*, Table II, if they were by measure equations only) to modern rational quantity (not measure) equation using rational, consistent units (Procedures *c–d*).

2. What is the oersted?

With "measure" equations this question is of little importance, provided only the measures of H and B are compared. The equation: 1 g = 1 oe, does not mean equal quantities $H=1\ \mathrm{oe}\equiv B=1\ \mathrm{g}$ but equal measures $H\equiv B$. Usually by "oersted," the unit 10 amp/4π cm (people's oersted) is meant, as the rationalization of units is usually understood to be an identification $\mathrm{H}'[H']\equiv\mathrm{H}[H]$ despite the difference in definition of the abstract quantities H' and H. So the people's oersted, from the quantity point of view, turns out to be the "people's gauss" instead.

If, on the contrary, we would like to answer the two questions from the quantity calculus point of view, it can be seen that most of the preceding classical procedure is realized in the transition from "quantity" rationalization Procedures *e* to *d*, Table II. In *e* the field of 1 g is realized by 10 amp-t/4πcm and $\mu=1$, as for the old-timer. But the magnetic field intensity "realized" simultaneously with $B_0=1$ g is 1 rational IEC oersted (10 amp/cm) as unit of H'' and not 1 nonrational people's oersted. The latter is "realized" in Procedure *a* as unit of H', but gives a different induction, namely $B_0=1/4\pi$ g, with $\mu_0=1$ also. But this 1 is more exactly 1 g/4πoe$_\mathrm{n}$. All this is further illustrated in Table I for $B_0=1$ g in standard solenoid.

The IEC Technical Committee 24 once tried to satisfy all sides with the following recommendation:

> "The equation between *units* is 1 amp/m = 10^{-3} oe. But the equation between *standards* (!) for these units is 1 amp/m = $4\pi\cdot 10^{-3}$ oe.
>
> "Oersted may be used with nonrational equations only, and amp/m with rational ones only in order to avoid confusion."

Old Pythia or the Delphi oracle could not have spoken more wisely

The solution, permitting one to avoid all these difficulties, may be pointed out in the following conclusions:

1. Instead of using equation 1 oe = 10 amp/4π cm or 10 amp-t/4π cm, say: (the induction $B_0=$) 1 g is produced by (the linear excitation θ) 10 amp-t/4π cm $\approx$ 0.8 amp-t/cm = 80 amp-t/m. Written in form of correspondence (or equation between measures):

$$1\ \mathrm{g}\triangleq 0.8\ \mathrm{amp\text{-}t/cm}$$

2. *In vacuo* (and also air practically) this (induction $B_0=$) 1 g corresponds to the nonrational magnetic field intensity 1 oe if oersted is understood in the old people's manner and equations come be-

Vol.
Pt. 1
1959

tween measures. Written again as correspondence for measures:

$1\ \text{g} \mathrel{\hat{=}} 1\ \text{oe}_{\text{n}}$

Be careful if using quantity equations for H!

3. The nonrational μ_0 of free space in this case is to be taken equal to unity ($\mu_0 = 1$).

4. The units amp/4π m, (mksa oersted), abamp/4π cm (the people's oersted), amp/m and abamp/cm (the IEC oersted), have a nonequivocal meaning and nonvarying physical effect (B_0) only if they are understood as units of the linear excitation and not as units of the magnetic field intensity H. It is better to add to them the physically illustrative "turn (t)." Units without t might be used with H only, *cum grano salis* (with a grain of salt). Ampere-turn per meter has a physical sense invoking immediately the solenoid experiment, which cannot be said of the ampere per meter.

5. The best way to get rid of these difficulties is to exclude their provocator, the hypothetical quantity magnetic field intensity H, by plotting B against θ, or, better still, against B_0 as B/B_0 gives directly the relative permeability. Many engineers do so already, the author included.

6. For the present transition period, it is highly recommended that engineers become acquainted with the problems as literature is written quite uncritically. The best way to learn the procedure is to calculate complete practical examples in all possible ways: measure and quantity, cgs electromagnetic, electrostatic and Gaussian, mks and mksa, rational and nonrational, and unit and quantity rationalization. It is the hard way, but with practice it is lots of fun and a much clearer insight into the meaning and content of physical equations of all kinds is obtained.

As seen, no high philosophy concerning electricity was used in the reasoning. But practical computations are unavoidable for proper understanding. "*Probieren geht über studieren*" (One single test [or standard] is worth a thousand expert opinions.)

The old way is simple, theoretically illogical but practically good, and the results are correct since two "false" steps cancel out.

To the young generation let us explain metrological problems correctly and systematically, especially in textbooks. Thus they will encounter less trouble in reading electrotechnical and electrophysical literature which is often not clear regarding equations, quantities, units, and measures, to the dismay of the uninitiated readers. Having had proper instruction, however, they will enjoy it and their work will be greatly facilitated. This is particularly true in magnetics where some veteran high school, and even college, teachers with outworn concepts slide over magnetism as fast as they can, without properly understanding it.

Only so the sleeping dogs will lie until they die.

Microwaves and Their Use in Power Systems

S. C. BARTLETT
MEMBER AIEE

MUCH HAS BEEN written on the subject of microwave, including theoretical considerations, characteristics, design of components, practical systems, experience, etc. Now that microwave systems have been in use for a period of years, it would seem of value to reappraise the situation from the viewpoint of the power system operator and engineer and to determine just what the true microwave characteristics are; to discuss relative costs, applications, reliability, and maintenance problems: in general, to consider how microwave can properly take its place to help carry the burden of system expansion. With such a purpose in mind the author has set down as briefly as possible those factors which have been observed in operating systems. Technical details have already been competently covered by such articles as those in the list of references; in this paper comments are confined to observed results and how well they substantiate theory.

The services rendered to the American public by the nation's electric utilities are fundamental to the existence of the American economy as we know it today; they enter every significant industrial process, and in the case of some are a major ingredient. Their benefits extend to farms, commercial establishments, and homes. Our economy with its high living standard would not have been possible without adequate and low cost supplies of electric energy. Total consumption has for many years been increasing at a rate beyond that of gross national production and there is no visible sign that this trend will not continue.

To provide all of this in the most efficient manner, larger and more efficient turbogenerators using higher steam pressures have been developed and built into vast network systems operating at ultrahigh voltage and interconnected on essentially a nation-wide basis. With this arrangement, plants may be strategically located to take maximum advantage of low cost fuel and by means of an efficient transmission system can convey energy to the load centers.

Simply the fact that there is a physical separation betweeen the dispatchers, plants, substations, interconnection points, and loads means that communications are called for in their broad sense, i.e., the transmission of intelligence in all forms from one point to another. The amount, quality, speed, and reliability of these communication systems are related to many factors of the power system, particularly its over-all size; the location of plants, lines, interconnecting points; the character of the machines and governors; and the character of the load which the customers apply. Basically a means of controlling generation is required, so that each system continually carries its own load by the most efficient means after having determined this load through observing the flow at all interconnections and adjusting generation accordingly. This involves a fast transfer of intelligence, since there may be rapid and frequent shifts of large loads and the net flow must be derived from concurrent observations. Here an automatic load controller can determine the corrective value required and transmit this information to the generating plants. The performance of such a system when provided with high-grade information channels is truly impressive and promises practical operation of a system with a margin of spinning reserve of less than 3%.

Operating the American Electric Power (AEP) System

In the AEP Service Corporation system the several operating companies acting in concert have accepted as obvious the

---

Paper 59-1099, recommended by the AIEE Power System Communications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted January 15, 1959; made available for printing June 24, 1959.

S. C. BARTLETT is with the American Electric Power Service Corporation, New York, N. Y.

thinking that superior over-all efficiency can be attained if the system as a whole is considered and regulated for maximum over-all advantage, and if all the factors just mentioned are considered. To do this involved establishing a load-coordinating office near the geographic center of the system and equipping it with all facilities for load forecasting including a penalty factory computer to accomplish this purpose. For the work of this office the various readings which must be delivered to it are shown on Fig. 1; some of these are interconnections and some are indications of plant net generation, but all require the same treatment as far as communication channels are concerned. In general these signals are used by the individual operating companies, being transmitted by power line carrier, leased lines, etc., and must now be picked up or extended to the Columbus headquarters. These signals are either of the pulse-duration or the variable-audiofrequency type.

The Problem

Telemetering, though essential to system operation, is by no means the whole solution. Information must be gathered by other means, particularly Teletype and facsimile. Voice, though it is a slow means for the transmission of intelligence, has its advantages and presumably will always be used. Automatic data processing is finding new uses every day and could be employed to analyze the values of interchange power flow, net generation, fuel costs, machine efficiencies, and line loss to determine the most efficient combination of generating units and the loading on each.

In the operation of the physical plant, relaying, transfer trip, supervisory control, alarms, and remote control, in addition to voice, require means of transferring intelligence from point to point.

For the activities of administrative and nonoperating personnel, communication is required primarily by telephone and teletypewriter. Fig. 2 shows the various headquarters, division and district offices, all equipped with private branch exchanges and some with private automatic exchanges, among which the majority of the traffic flows. The loading by leased wires and long distance calls between each pair of these points must be added in, to determine which may justify either private or party lines.

The question then boils down simply to: "With these requirements, what is the most economical means of providing all of these circuits with suitable reliability, keeping in mind expected growth rates?"

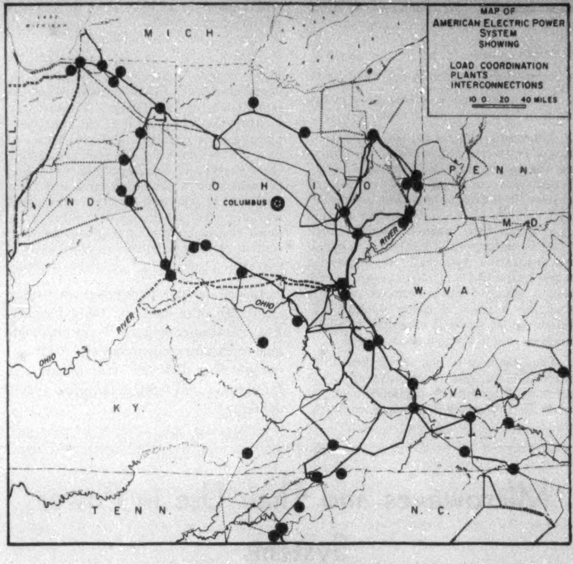


Fig. 1. Telemetering terminals

Factors Considered

A detailed study was made to determine the rate of growth of communication facilities based on channel miles from which estimates were made for the requirements for each of the next 10 years.

Table I gives the figures on which this estimate was based, and shows that 10.1% is the average annual increase for all types of facilities. In this estimate mobile radio which serves an entirely separate function from the point-to-point services has been intentionally excluded. It will be noted that the rate of growth indicated above is greater than for the power system itself, which is about 8%. This, it is felt, is realistic because of the additional functions being assumed by the communication systems. Thus appears the problem from the indicated rate of increase in channel miles, and there seems to be no indication that this trend is changing.

The next step was that of setting down the locations of all channel terminals and deciding which might be grouped or brought to assembly points to be handled by some common means such as a microwave system. The spelling out of the problem in terms of specific channels, types of intelligence to be conveyed, terminal locations, operating dates, rate of growth, etc., is complex but must be done. It would avail nothing to go into these details here since no two power systems bear more than a superficial resemblance to each other. The next step was to consider all possible tools and their relative costs. All practical means having been considered, it was found that the problem boiled down rapidly to the use of the following, either separately or in combination:

1. Power line carrier.
2. Leased wires.
3. Private telephone line.
4. Carrier on private telephone line.
5. Microwave.

To explore the economics, which are always of paramount importance, it is necessary to arrive at an evaluation of the annual costs for capital expenditures to balance against the cost of leased facilities. In the several systems considered the various components have different expected lives but the weighted average for the cost of each component and its expected life proves to be around 20 years for each entire system.

Cost Figures

Although the cost of money varies with the proportions derived from bonds, pre-

Vol.
Pt. 1
195

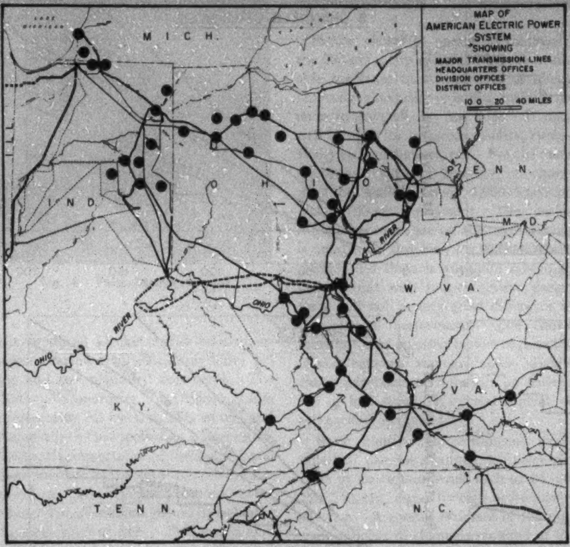


Fig. 2. System offices

ferred stock, and common stock, and the interest rates on each, it is necessary to have at least an estimating figure for engineering purposes. At the moment, these figures plus income tax on preferred and common stocks plus local property taxes, sinking fund to redeem in 20 years, and maintenance, all add to almost exactly 20% per annum on the total investment.

Making a number of assumptions, particularly in estimating the cost of building the least expensive systems, such as transmission telephone systems on wood poles along existing rights of way using long-span rural construction, phantoms wherever possible, repeaters where necessary to maintain a suitable signal-to-noise ratio on circuits with high exposure to power line noise, resulted in the series of curves of Fig. 3. The estimate for leased facilities is based on a cost of $4.00 per mile per month plus 10% tax for voice circuits with no concession in the way of a sliding scale either as a function of length or number of circuits in parallel. For a microwave system, average figures have been taken for terminal stations and repeater stations, assuming 25 miles each for the hops.

These curves indicate that power line carrier is in general the least expensive and yet, valuable as it has been to the AEP through the years, when it has been exploited to the fullest, it must be recognized that there are very definite limitations, particularly in spectrum and in practical length, the latter being restricted by the manner in which high-tension systems today are coming to look more and more like distribution networks. New techniques, circuits, tubes, and components have helped somewhat in the problem of adding channels but the limits are rigid, particularly because of the side-band requirements of voice and relay channels so that, at best, one may expect only a small percentage of additional channels to be handled by carrier. Economics dictates that power line carrier be used to the limit of its technical capabilities, particularly for short haul, feeders, etc., but other means must be found for handling the majority of channels.

Leased telephone circuits serve a valuable purpose in this business field, being available for almost any bandwidth requirements and in any number of circuits. However, they have the inherent weakness of any open-wire or cable system aggravated to some extent, probably by the fact that the AEP plants and substations are frequently in areas remote from telephone operating centers. The curves of Fig. 3 show that on a cost basis their use is very definitely in the picture, particularly where a small number of channels is required.

Table I. Requirement Estimate Basis

Type of Communication	Channel Miles 1945	Channel Miles 1954
Telephone	2,817	6,033
Relay	2,921	6,026
Telemeter	2,981	8,576
Totals	8,719	20,635

The average increment per year is 10.1%.

Private telephone systems, either metallic or by carrier, are technically no better than leased lines but may be preferred in that all equipment is under the control of the user. On the other hand, outages on leased wires may be minimized by having available circuits by alternate routing. Economically, the figures show that privately owned wire line facilities are never the solution except, possibly, in some cases requiring an unusually large number of channels in parallel.

Microwave fits in where comparatively few channels are required and its technical characteristics should be explored to determine suitability.

Carrier current being essentially saturated and private telephone systems economically unattractive, a solution must be reached between leased lines and microwave or, more likely, a combination of both. In cases where either would serve, it appears that microwave has the advantage on at least the following technical grounds:

1. Less outage time.
2. Concentration of all equipment at a small number of points.
3. Completely under user's control.
4. Ability to restore audio level at all points.

Of somewhat less importance and possibly debatable are:

1. Higher intelligibility.
2. Better signal-to-noise ratio.
3. Lower crosstalk level.

This leads to the conclusion that the true crossover between leased lines and microwave shoud be at a smaller number of channels than indicated on the basis of economic considerations only. The microwave characteristics will now be considered more in detail to determine if it is reasonable to commit many important signals to a system involving an r-f (radio-frequency) power comparable to that of of a flashlight bulb.

Microwave Licensing

In considering microwaves which in general are understood to be those fre-

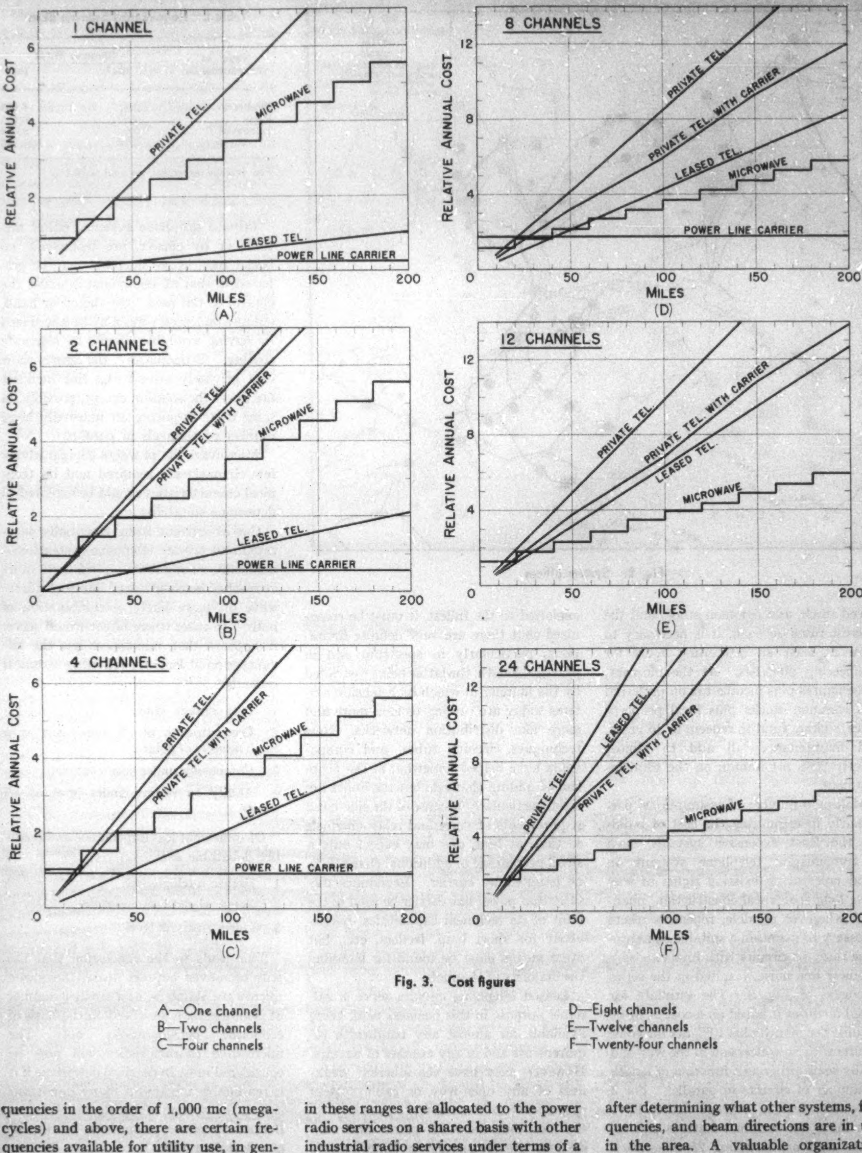


Fig. 3. Cost figures

A—One channel
B—Two channels
C—Four channels
D—Eight channels
E—Twelve channels
F—Twenty-four channels

quencies in the order of 1,000 mc (megacycles) and above, there are certain frequencies available for utility use, in general around 960, 1,900, 2,200, 2,600, and 6,700 mc, and consideration must be given to the propagation characteristics of these frequencies and to equipment available for operation here. Frequencies in these ranges are allocated to the power radio services on a shared basis with other industrial radio services under terms of a developmental grant only. This means that an applicant accepts the condition of a license which may be revoked at any time, that he must be aware of other users, and that he select operating frequencies after determining what other systems, frequencies, and beam directions are in use in the area. A valuable organization known as the Operational Fixed Microwave Council has been quite effective in tabulating all such information which is available to any potential utility licensee. The Federal Communications Commission

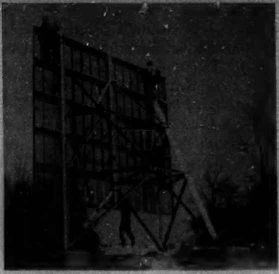


Fig. 4. Philo passive repeater

has produced Docket *11866* which considers a review of all allocations above 890 mc and has requested comments. The National Committee for Utilities Radio representing all electric, gas, water, and steam utilities has replied and, after establishing the importance of the utilities and their need for bands in the microwave frequencies, has suggested among other things that allocations should be in blocks for each class of user. This would simplify the problem of frequency allocation, particularly for the utilities, since each operates in an established franchise area. Obviously maximum utilization of the spectrum calls for careful consideration of all factors of propagation, channel loading, beam width, deviation, etc., and the objective of utility users is to establish their basic requirements and rate of expansion, and to co-operate in seeking the best solution for all concerned.

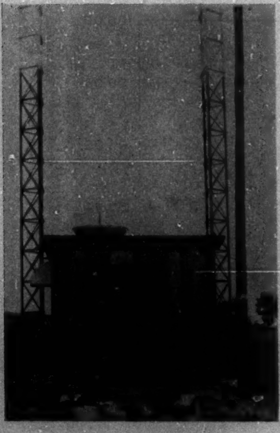


Fig. 5. Typical steel building

Systems Available

In the choice of operating frequency the 960-mc group is limited by the available spectrum to a maximum of about ten voice channels; it is also limited by noise to systems of a few hops. It may well be ideal for branches, feeders, supervisory control systems, etc. For the main microwave system, however, a choice must be made between 1,900–2,600-mc and 6,700-mc equipment. The necessary components are on the market from a number of suppliers for both of these ranges, and there is adequate reason to believe that with proper system design there will not be any great difference in the over-all performance. The choice must be based on other factors such as the choice between waveguides or coaxial cable, parabolas or reflectors on towers, full or half stand-by, crystal control, or self-excited oscillators, etc.

Basic multiplexing equipment available will provide channels for voice circuits up to certain levels such as 24, 60, 120, 240, etc., either by time or frequency division. A coreful estimate of the ultimate number of channels required at the time of amortizing the equipment must be made in order not to carry a burden of unnecessary basic equipment or to run into situations where a major change must be made in the basic multiplex equipment. If there is considerable need for drop-and-insert channels, frequency division probably has the advantage. On the other hand, long through channels may dictate time division to attain a better signal-to-noise ratio.

Although the economic and technical factors mentioned above justify full consideration of microwave, it must be realized that the high initial cost limits its use to cases where either many signals are required or to areas where lines are impractical, or where extremely wide bands such as for television must be conveyed, or for temporary installations where it would be uneconomical to build lines. In the utility business this brings up something of a new concept of investment in communication equipment: the purchase of coupling capacitors, line traps, and carrier current equipment would involve only a few thousand dollars each, whereas a microwave system involves a long-term investment which, even for a simple system, may run into six figures.

First AEP Installation

For a number of years the AEP system studied the technical characteristics and particularly the performance of microwave systems and, in fact, observed demonstrations of 960- and 4,000-mc equipment in operation during 1947. However, an extensive power line carrier system made it possible at low cost to add channels from time to time, so that it was not until 1953 that carrier saturation forced a decision regarding alternate means, and the engineering studies singled out the use of microwave as the clear-cut answer to most of the problems. By this time, microwave equipment and associated multiplexing devices had reached a relatively suitable stage of development, many of the early problems having been solved, and components improved, so that there were few misgivings. Then AEP proceeded with the first installation extending some 80 miles from Canton (Ohio power headquarters) to Philo (major generating plant) using three repeaters. The only unique feature of this first system is the large passive repeater, constructed entirely of aluminum, which eliminates a full repeater station or a long telephone cable line to get the signals into final plant about a mile away across the Muskingum River; see Fig. 4. This passive repeater, more than any other factor, dictated the use of 7,000 mc. Six voice circuits were provided, one of them being subdivided by audio tones for telemetering. Figs. 5 and 6 show a typical repeater station and its antenna system. Fig. 7 is interior of a terminal station.

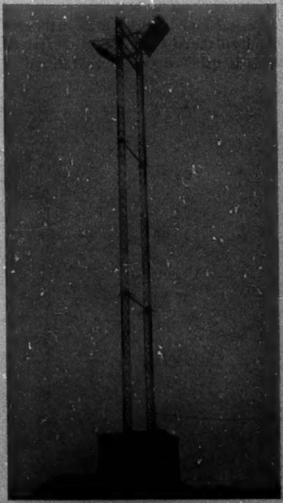


Fig. 6. Repeater with tower

Fig. 7. Terminal station interior

Results

By the end of the first year, in spite of the original errors, inexperience, and those defects which are expected to show up in the first few months, it was obvious that the system fulfilled all predictions: the cost was as estimated, the technical performance was as claimed, maintenance was less than expected, and continuity well above the guarantee.

Reliability may properly be defined in many ways, depending upon the function of the equipment, but for purposes of appraisal it is interpreted to mean the continuity of the r-f signal through the system. This is qualified to mean r-f signal of sufficient magnitude to deliver understandable signals, signal/noise = 15 db (decibels) on any voice channel. Experience shows, however, that this detail is unnecessary since equipment failures or fades seldom produce losses approaching path margins: they usually cause either a complete outage or nothing. Fig. 8 gives a record of performance extended to the present, showing continuity above 99.88% for the system or 99.97% per hop.

Microwave was thus accepted as a proper tool for the solution of problems of channel overloading elesewhere in the system. In other areas the shortage of channels was becoming acute, and restraint was required to avoid a sporadic growth of scattered microwave systems. Thus, before proceeding, it was necessary to develop a long range plan so that, as these systems were constructed they would grow together and co-ordinate in an orderly fashion. A 10-year plan was formulated, as illustrated in Fig. 9. The AEP was considered as a whole, including rate of growth, locations of generating plants, substations, loads, business offices, interconnections, etc., and the best possible estimate was made as to the location of new lines, plants, and loads.

The original Canton-to-Philo system of 1954 was followed by an extension westward to Newark (accounting headquarters) in 1956, then to Columbus (load dispatching office) in 1957; also an extension southward to Muskingum (the major plant) was made in 1956, and to Sporn (another major plant) in 1957.

Similarly, problems of channel congestion appeared in the Indiana and Michigan Electric Company which could now be solved by similar means. Here, 1,900-mc equipment was chosen, primarily on the basis that two of the five locations already had radio towers which would serve to support parabolas but would not be capable of handling the wind loading of the considerably larger reflectors associated with 6,700-mc equipment. Also, largely for comparison purposes, single side-band multiplexing was chosen, whereas frequency modulation was used in the Ohio system. The initial microwave system ran from Robison Park (load dispatching office) to Spy Run (company headquarters in Fort Wayne) to Sorenson (345-kv substation) and Deer Creek (major substation and division headquarters) in 1957.

Similarly, in the Appalachian Power Company a 145-mile system was built between Roanoke, Va. (headquarters office), and Charleston, W. Va. (division headquarters). Here again the most desirable path involved the use of a passive repeater. In this case it could be mounted on a water tank not owned by the AEP, but only if its size and weight could be held to a minimum, so that again 6,700-mc equipment was chosen; see Fig. 10. This system is unique in that no towers are used at any of its eight stations, since six of them are on mountain peaks and the terminals on office buildings (with the help of the passive repeater just mentioned). Fig. 11 shows a typical mountain top repeater station.

Fig. 9 indicates by solid lines the portion of the AEP system which is now in operation; this portion constitutes the basic backbone of the system, and it closes the gaps all the way from South Bend,

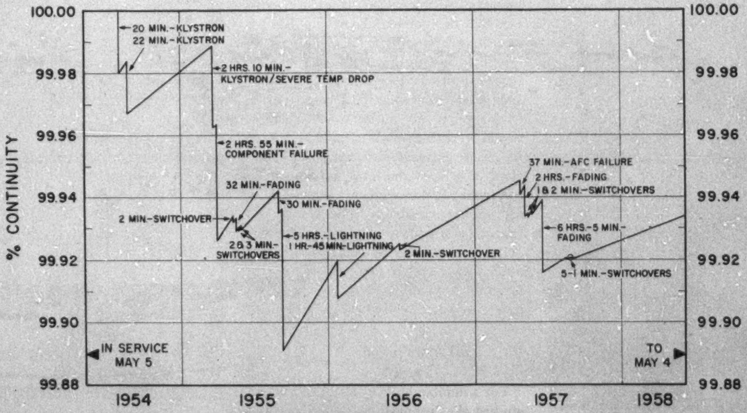


Fig. 8. Continuity of Canton-Philo microwave

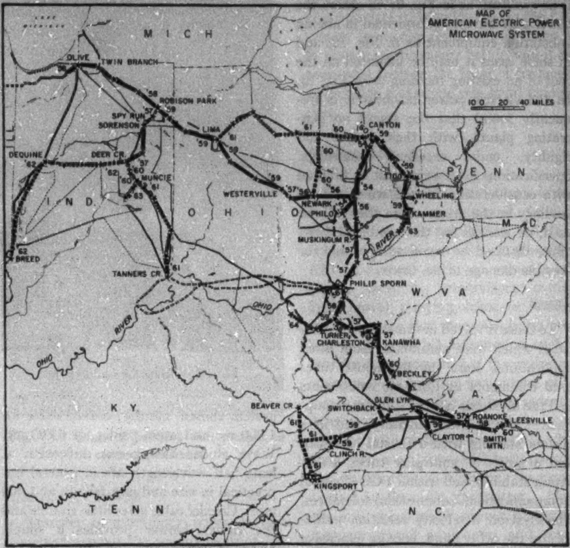

MAP OF
AMERICAN ELECTRIC POWER
MICROWAVE SYSTEM
10 0 20 40 MILES
MICH
OLIVE
TWIN BRANCH
ROBISON PARK
SPY RUN
SORENSON
LIMA
DEQUINE
DEER CR.
MUNCIE
IND.
OHIO
WESTERVILLE
NEWARK
PHILO
MUSKINGUM R.
CANTON
TIDD
WHEELING
KAMMER
PENN
M. D.
RIVER
BREED
TANNERS CR.
PHILIP SPORN
W. VA
OHIO
RIVER
OHIO
TURNER
CHARLESTON
KANAWHA
BECKLEY
VA.
K Y.
BEAVER CR.
SWITCHBACK
GLEN LYN
ROANOKE
LEESVILLE
CLAYTOR
SMITH MTN.
CLINCH R.
KINGSPORT
TENN
N C.

Fig. 9. Map indicating the 10-year plan

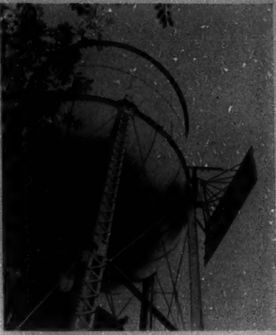


Fig. 10. Passive repeater on water tank

Fig. 11. Typical concrete building on mountain top

Ind., to Roanoke, Va. Although there may not be a through channel from end to end of the system, the design must be adequate to permit one or more talking circuits from Fort Wayne to Roanoke, probably by an audio patch at Columbus.

From here on, additions will be made as economic crossovers are reached, more or less along the lines of the long-range plan, which will doubtless be altered to accommodate new lines, loads, and plants which are certain to materialize.

Observations

Theoretical considerations have been adequately covered by various qualified writers; yet predictions contain differences of opinion, and the question always arises as to which of these will prove valid in practice. The manufacturer in establishing the technical advantages of his product may be expected to emphasize its good characteristics. The user is left to evaluate the importance of each characteristic in his particular application. Thus it should be of value to appraise these systems on a basis of experience and to indicate which of the characteristics are really important in the operation of electric power systems, which ones stand up in practice, and which ones may still be of debatable value because of small differences.

FREQUENCY

The proponents of 1,900- and 6,700-mc systems continue to debate as to which of these produces the less total product of depth of fade and duration; our observation is that with average good design and comparable cost either one can attain the 99.97% continuity per hop mentioned above. The argument is thus largely academic.

CHOICE OF PATH

When the operating frequency has been selected, a suitable path may be chosen. The AEP method is to work from United States Geological Survey maps, drawing tentative profiles providing 0.6 Fresnel zone plus 50 feet provided arbitrarily for trees and other obstructions. A number of paths may be plotted in seeking the best compromise between tower height and path length. Also, care must be taken to choose locations as close as possible to roads, power lines, etc. With several possible paths the next step is a field check of the sites, some of which may be found unavailable due to prior occupancy, unsuitable ground condition, difficult access or, as has been encountered the site is, in effect, missing as a result of strip mining or earth-moving operations.

The theory is well established regarding the relation between path loss and Fresnel zone, and show that there is little to be gained in going beyond 0.6 of the first Fresnel zone. If a few db are important it is better to obtain them through the use of larger parabolas. Less well established is the basic earth curvature to be used. From this company's limited experience it would appear that about equal performance results with a 1,900-mc system based on a 4/3 earth curvature or from a 6,700-mc system based on true earth curvature, each being designed for 0.6 of the first Fresnel zone.

The next step is to check by survey either on foot, with altimaters and flashing lights or by means of an aerial survey. Each has its own advantages, e.g., the survey on foot is most assuring, particularly if actual light flashing can be accomplished before the towers are built. Also, in cities, ground survey is necessary since low-flying planes are prohibited. On the other hand, radar profiling is particularly good in rough country and where dense growths are encountered There may also be some merit in obtaining the necessary information without disclosing plans to property owners.

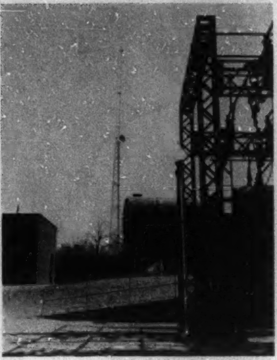


Fig. 12. Tower in station yard

ANTENNAS

Antenna gain defined as the ratio of field strength in a given direction to the field strength of an isotropic antenna, although running in the thousands, is realistic, as well as are the relations between parabola size and beams widths.

PASSIVE REPEATERS

In the use of passive repeaters, it is found that both those used at the tops of towers and those of the billboard type give very satisfactory performance. The losses with properly designed sizes and terminal parabolas based on the lengths of the two path legs are negligible and, in fact, may be slightly negative due to some focusing effect. A large-sized passive repeater such as the one mentioned above for Philo Plant, Fig. 4, costs about $10,000 installed and yet is far less expensive than an active repeater station, and certainly better than a wire line or cable for the short leg. It is felt that many applications will appear, particularly in mountainous country where at the end of a series of relay stations a short hop is required to get down into the valley to the station. These may also serve in cities for bending a beam horizontally around buildings or other tall structures; see Fig. 10.

TOWERS

When it is necessary to consider the supports for parabolas or reflectors, the choice is not hard to make. Single wood or H frames for heights up to 70 feet, and steel towers for heights above that level are entirely satisfactory. If the available area allows, guyed towers permit some saving in cost, otherwise self-supporting towers must be used. In some areas space is no problem, as in Fig. 6; in others the tower may have to be squeezed in among substation equipment as in Fig. 12, and in some cases it may be mounted on the roof of an existing building as in Fig. 13. In the electric power business it is frequently found that the stacks on generating plants, with their height and rigidity, make excellent mountings. Specifications for towers must be tailored with consideration for the wind and ice loadings in the areas where they will be used, as well as for the bend and twist limits dictated by beam widths, and for possible damage to the tower.

TUBES

To those involved in this study the only tubes which were unfamiliar, from other applications, were the lighthouse tubes used commonly in 1,900-mc equipments and the klystrons used in 6,700-mc equipment. The former, generally used in a system starting with a crystal oscillator and a series of multiplier tubes, easily attain stability well within FCC (Federal Communications Commission) tolerances, The klystron, a velocity variation oscillator, on the other hand, is more expensive and requires precision circuit design to attain the necessary stability but it is in itself the oscillator and final tube; thus its use avoids the series of multipliers. Failure of its oven or thermostat control could permit a drift beyond operating limits. Obviously nothing but tubes of the highest quality can be considered in this service and they should be subjected to preliminary aging of at least 48 hours before going into service. The spare r-f assembly, if operated hot, is an excellent place to run in these new tubes.

BUILDINGS

In the buildings both steel and concrete are used; see Figs. 5 and 11. Both have been found satisfactory and with not very much cost differential. Steel is readily assembled and easily expandable but may require special caulking and insulation. For isolated locations concrete affords some additional protection, particularly from bullets. For 6,700 mc, where antennas are normally mounted on the roof and pointed upward, it is a simple matter with the concrete building to extend the wall to form a parapet above the roof, concealing the parabolas; see Fig. 14. Naturally an existing building would be used wherever possible. Equipment could be arranged as in Fig. 15.

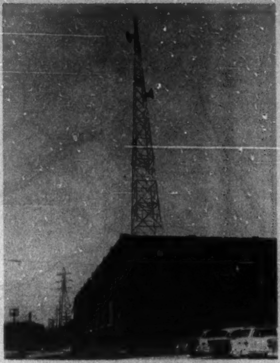


Fig. 13. Tower on building roof

R-F LINES

As far as the r-f feed system is concerned, it appears to be tied rigidly to the system chosen, namely, coaxial cable for 1,900 mc and wave guides for 6,700 mc. Wave guides having one dimension at least half wavelength are in general impractical in size and cost for use at 1,900 mc. Coaxial cable gives little trouble and the cavity blower provides a simple means of pressurizing the line. At 6,700 mc the wave guide and its low loss are ideal and universally employed. By putting the parabola at or near the transmitter there is practically no run of r-f line and no problem of maintaining cable on towers. Wave guides, being frequency selective, and having a cutoff frequency at about half operating frequency, are sometimes cited as an advantage over the coaxial cable. The ability to provide selection so far away in frequency does not make an impressive argument.

STANDBY R-F

In the matter of standby r-f equipment the individual application will dictate the rate of transfer, which in the case of relaying must be in milliseconds to those cases where only voice channels are handled and even an interruption of several minutes may be tolerated. To minimize interruptions during transfer to spare r-f equipment, hot standby is required along with fast-acting wave guide or coaxial switches or, for complete continuity both transmitters may be fed continuously to separate antennas. Those systems using only two frequencies may employ a single set of spare equipment at a repeater station, cutting it in for either direction as required. Since a simultaneous failure of both normal equipments at a repeater station is a remote possibility, there seems no error in judgment to use the so-called one-half stand-by.

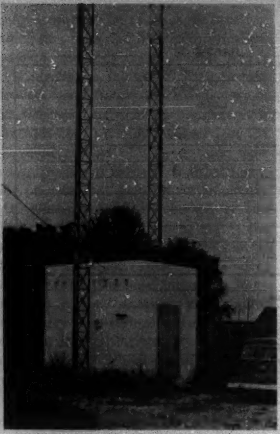


Fig. 14. Concrete building with parapet to conceal parabolas

EMERGENCY POWER

Transfer to emergency power is also designed around the requirements of the particular services rendered, and may be anything from a completely continuous supply to intentional provision of time for the line reclosers to function before starting the emergency generator. These machines give remarkable performance, particularly those operating on propane gas where heaters are provided either in the crankcase or the cooling jacket. In a matter of 1 or 2 seconds they will come up to speed and are immediately capable of picking up full rated load.

For relaying applications where outages must be strictly limited, particularly those at the instant of power supply failure, there are several approaches, all of them practical. Although diversity transmitters are not approved by the FCC it is possible to get good protection against fading by receivers spaced less than one mile. To eliminate even the momentary breaks during transfer to emergency power either battery-operated r-f or flywheel generator sets are practical.

MULTIPLES

The choice of multiplex equipment must be made between frequency division and one of the several forms of time division. Many factors will influence this choice, particularly the number of party lines, the over-all length of the system, the number of hops in a single voice path, the number of branch stations, the number of drop repeaters, and, where relaying is used, the speed of response. Unquestionably either system will, in general, give complete satisfaction, but in the AEP system frequency division was found to fit system requirements better, and it has been used both in the form of frequency modulation and single side-band modulation.

Applications

The versatility of these systems is such that with any conventional form of multiplexing it is possible to handle all forms of signals normally employed in AEP business, such as voice, telegraph, facsimile, supervisory control, load control, relaying, telemeter, alarms, etc. It seems that there will be no problem to fit in coded information such as that for data processing or the indications which may be required for systems of the future which have not yet been invented. Microwave also makes a convenient combination with mobile radio systems where the towers may be used for both purposes and the microwave system as the remote control for the radio transmitter. About the only thing which cannot be handled without special equipment and special developmental grant from the FCC is a television signal.

Although in the AEP, relaying and transfer trip are still handled satisfactorily by power line carrier, convincing indications are that no technical problem of any magnitude exists in handling these signals by microwave. The matter of obtaining complete continuity is readily solved by one of the methods mentioned above. In practically every case a microwave system must have duplicate r-f equipment and, if it is to handle relay signals, about the only change necessary is that this spare r-f must be arranged to run continuously. This involves comparatively little cost and has the additional advantage that it may be continuously monitored, giving the assurance that a transfer will never be made to a faulty spare unit. This matter of continuous supervision is itself an advantage in relaying since in general there is no continuously running signal on power line carrier and no means of determining that the channel is actually ready to function. A number of technical papers, included in the list of references, cover this subject.

Fading

Most factors in the design of microwave systems such as transmitter power, path loss, cable loss, antenna gain, and receiver sensitivity are readily measured. The exception is the factor of fading due to variations in atmospheric conditions, which must be considered largely on a statistical basis and weighed if possible by experience with other systems in the area. Although it might be theoretically possible to design a system where all fades are within the operating margin, the practical solution usually involves providing something in the order of 30 db per hop margin, which in turn produces an outage due to propagation failure of less than 0.02%. As far as one may observe, there are no reflections from the ionosphere sufficient to affect performance, and space noise from either atmospheric or man-made interference is below that of the front end thermal noise of receivers. It is not known whether fading is due to multipath propagation, refraction, atmospheric attenuation, or ducting, but each appears to have its own effect. For instance, slight signal loss in heavy rain or snow, definite loss in fog, rapid, and severe signal variations for an hour or more for no known reason, and fades for 20 minutes at a time so deep that there was no measurable grid current in the final limiter have been observed. Fortunately, only the last-mentioned case caused system outage and whole years have passed with no outage due to fading.

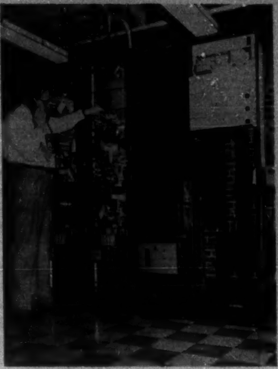
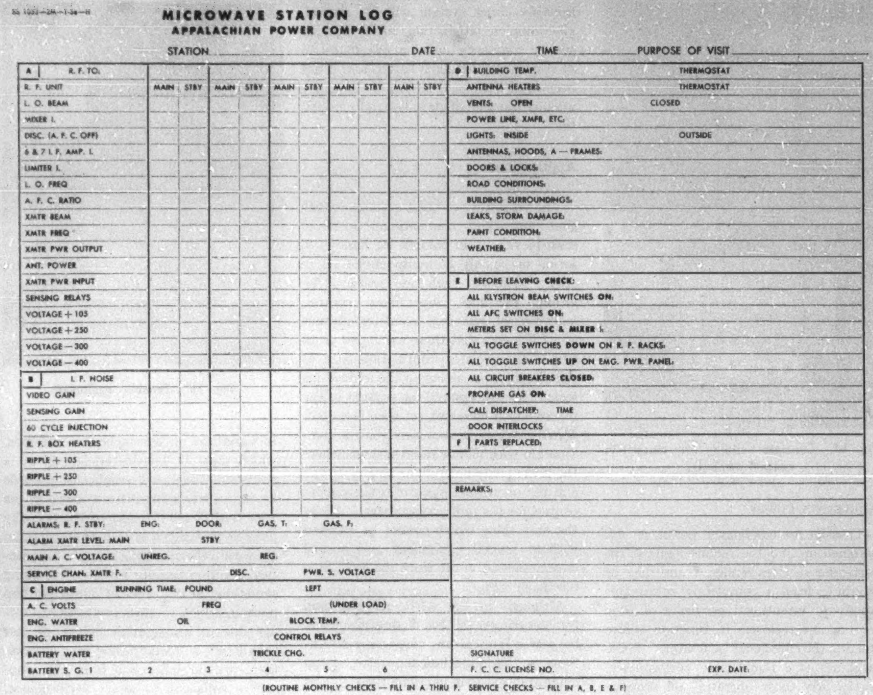


Fig. 15. Terminal equipment

Another interesting point is that such outages have always happened in paths of 20 to 25 miles in length, longer and shorter paths being affected. Also interesting is the observation that on one occasion the system was interrupted by simultaneous deep fades in two adjacent hops.

Maintenance

Routine maintenance practices must be adequate to hold systems constantly



SS 1053—2M—1-58—H

MICROWAVE STATION LOG
APPALACHIAN POWER COMPANY

STATION ____________ DATE ________ TIME ________ PURPOSE OF VISIT ____________

A R. F. TO:										
R. F. UNIT	MAIN	STBY	MAIN	STBY	MAIN	STBY	MAIN	STBY	MAIN	STBY
L. O. BEAM										
MIXER I.										
DISC. (A. F. C. OFF)										
6 & 7 I. F. AMP. I.										
LIMITER I.										
L. O. FREQ										
A. F. C. RATIO										
XMTR BEAM										
XMTR FREQ										
XMTR PWR OUTPUT										
ANT. POWER										
XMTR PWR INPUT										
SENSING RELAYS										
VOLTAGE + 105										
VOLTAGE + 250										
VOLTAGE — 300										
VOLTAGE — 400										
B I. F. NOISE										
VIDEO GAIN										
SENSING GAIN										
60 CYCLE INJECTION										
R. F. BOX HEATERS										
RIPPLE + 105										
RIPPLE + 250										
RIPPLE — 300										
RIPPLE — 400										

ALARMS: R. F. STBY: ENG: DOOR: GAS. T: GAS. F:

ALARM XMTR LEVEL: MAIN STBY

MAIN A. C. VOLTAGE: UNREG. REG.

SERVICE CHAN: XMTR F. DISC. PWR. S. VOLTAGE

C ENGINE RUNNING TIME: FOUND LEFT

A. C. VOLTS FREQ (UNDER LOAD)

ENG. WATER OIL BLOCK TEMP.

ENG. ANTIFREEZE CONTROL RELAYS

BATTERY WATER TRICKLE CHG.

BATTERY S. G. 1 2 3 4 5 6

D BUILDING TEMP. THERMOSTAT

ANTENNA HEATERS THERMOSTAT

VENTS: OPEN CLOSED

POWER LINE, XMFR, ETC.

LIGHTS: INSIDE OUTSIDE

ANTENNAS, HOODS, A — FRAMES:

DOORS & LOCKS:

ROAD CONDITIONS:

BUILDING SURROUNDINGS:

LEAKS, STORM DAMAGE:

PAINT CONDITION:

WEATHER:

E BEFORE LEAVING **CHECK:**

ALL KLYSTRON BEAM SWITCHES **ON:**

ALL AFC SWITCHES **ON:**

METERS SET ON **DISC & MIXER** I:

ALL TOGGLE SWITCHES **DOWN** ON R. F. RACKS:

ALL TOGGLE SWITCHES **UP** ON EMG. PWR. PANEL:

ALL CIRCUIT BREAKERS **CLOSED:**

PROPANE GAS **ON:**

CALL DISPATCHER: TIME

DOOR INTERLOCKS

F PARTS REPLACED:

REMARKS:

SIGNATURE

F. C. C. LICENSE NO. EXP. DATE:

(ROUTINE MONTHLY CHECKS — FILL IN A THRU F. SERVICE CHECKS — FILL IN A, B, E & F)

Fig. 16. Station log

within a few db of normal of design figures. In general it is found that, although one man can maintain about ten stations, most of his time is spent in traveling and it is more efficient to have a number of trained men available along the system for part-time work. By having men closer to the stations both outage time and traveling time are reduced.

For this work, the cost of the equipment and importance of continuous service is such that only the best men should be chosen. All suppliers offer training courses which should be taken at least by the first group of men before the equipment is received so that they will be able to make the installation, tests, and adjustments as well as take care of subsequent maintenance.

Alarm systems should be designed to convey enough information for an intelligent decision as to calling out the maintenance man immediately, the next day, or the next normal working day. It is considered important to have at each station a record of normal current and voltage readings for that particular equipment. A blank form, such as that of Fig. 16, was made up for a particular type of equipment and is valuable in that the maintenance man is frequently able to spot the trouble in a minute or two simply by observing deviations from previous readings. Section *E* on this chart is important for a check on the serviceman himself since it is not unknown for a serviceman, after completing his work, to walk off leaving some item in an inoperative condition.

On the interesting and important question of maintenance costs, a number of factors are involved, yet for all practical purposes one may reach a representative figure by imagining a typical system, say, 200 miles along with average hops and multiplexing. Such a system might use ten stations and cost about $265,000 installed. It could be maintained by the equivalent of one man, the annual cost of which, including his car, replacement parts, etc., might be $11,000, giving about 4% as an average figure. Field experience shows a cost slightly above this for the first year, and somewhat less subsequently as the equipment settles down.

Evaluation of Microwave

From field experience AEP observers have satisfied themselves that microwave has advantages over alternate means of communication for the purposes of utility operation, at least in matters of higher intelligibility, better signal-to-noise ratio, ability to restore zero db above 1 milliwatt level, lower crosstalk level, and better continuity. It also has flexibility by which it can be adapted to handling any of the many forms of intelligence required. Once the channel requirements exceed the order of six or eight voice circuits, the microwave also has economic advantages; equipment is concentrated at discrete points, which facilitates servicing. Theoretical factors stand up well

Vol. 7
Pt. 1
1959

in practice and details such as the causes of fading are, from the AEP point of view, academic only, since the over-all outage time results, as related to path margin derived from theory, are satisfactory in practice. Other definite tangible advantages not so readily evaluated are the increased number of long distance calls made because of inadequate other facilities or time lost waiting for a circuit. Maintenance costs per channel or per channel mile are no burden, nor is the matter of training personnel. In general the AEP feels secure in making microwave responsible for the most important functions of system operation; its apparent complexity is no serious matter since designs are made with great care to hold the number of components in the common equipment to an absolute minimum.

Microwave being normally a multichannel device may reasonably be expected to become a backbone system for long hauls to which other systems may act as local area feeders. Thus the microwave system should be engineered to high quality standards about as a toll telephone system would be, so that the signal deterioration in long hauls would be no greater than in local loops. Values in the order of 42 decibels adjusted (*F1A* weighting) should be obtainable in systems of ten repeaters without investing in more than moderate-priced equipment.

The AEP has here the basis, if desired, for setting up a full company-wide direct-dialing system by means of which it would be possible to dial any extension in any plant or office in the company. This could be done by using private automatic exchange switching machines which are now available in sizes down to those suitable for the smallest office, or by terminating on leased boards.

In the matter of terminating microwave voice circuits on leased telephone boards or continuing over leased wires, there is an agreement between the Bell Telephone Companies and right-of-way companies. The general effect of this agreement is that such connections will be made on the understanding that their use will be confined within the latter companies' right-of-way or that of neighbor companies with which it does business. The same general policy is followed by the independent telephone companies; it provides the protection which they require for their local and long-distance traffic, and leads to an increased use of leased wires and terminal facilities, which might not be justified if microwave were not available for part of the distance.

Party lines are by no means things of the past; in many cases they are valuable for alerting or transmitting information simultaneously to many substations. Also, it is logical to use the party line type of equipment for circuits to a number of unattended substations where the total traffic may readily be handled by a single circuit. Again, circuits having a simple ring-down are valuable where there is a need for direct and instantaneous contact between parties and no delays can be tolerated.

Channels of any type can be set up between any desired points, either throughout the entire system or only in local areas, and can be added singly up to the ultimate capacity. Control signals such as those used for teletype, telemeter, etc., require little channel space, and by a method of tones, 18 such signals can be applied simultaneoulsy over any voice channel.

In some cases ultimate channel requirements may be so high that consideration should be given to the maximum number of circuits which should be entrusted to any medium, no matter how nearly perfect. Although unlikely, it must still be realized that a system could be put of service for an extended period through loss of a tower or electronic equipment by tornado, fire, sabotage, or enemy action.

Conclusions

The dynamic nature of the industry must be recognized and one must be prepared to bring into use any system or method which holds promise of leading to improvements in safety and economy or the well being of the general public. Microwave should be given very thorough consideration in programs of expansions, not only as a means of handling channels for conventional purposes but for new applications, data processing, for instance and new inventions requiring the transmission of intelligence, as they may appear from time to time.

The present-day trend of seeking to obtain maximum operating and economic advantages leads to vast and complicated power networks, the efficient operation of which requires substantial increases in both the quantity and quality of communication channels. A factor in planning of considerable importance is the ultimate channel capacity required in any area. Power systems are expected to grow without limit, yet the microwave system must be designed for some ultimate channel capacity. Since the basic equipment is expensive, even though only the working multiplex is provided, it is necessary to choose a combination of microwave system and multiplexing equipment suitable for the channel loading anticipated at retirement age, say, 20 years. It is sometimes difficult to conceive of the present rate of system growths continuing indefinitely, yet this must be assumed. It is necessary to make the best possible estimate for growth in each area, designing communication systems accordingly, and if anything, leaning in the direction of making generous provisions. Although in utility work high-capacity communications systems are not generally required, it is important to watch carefully those areas of concentration such as at headquarters offices, load-dispatching centers, etc., where the microwave is originally justified; it is likely that traffic at these points will determine ultimate channel future requirements.

It is not intended that microwave should assume the full burden of communications but rather that it should take its place to share the burden with other methods, each being applied to take maximum advantage of its individual characteristics.

Ordinarily, new circuits which cannot be handled by carrier current would logically be applied to leased lines, and as the number of these circuits increased to the crossover point of 6 or 8 voice circuits in parallel, it is expected that they would then be replaced by microwave. Once the microwave is installed, it may then be used to shorten those carrier current channels which are too long, giving an over-all improvement in signal-to-noise ratio. Many new applications will appear through having additional circuits available at small cost. The AEP has in mind, for instance, management circuits either from headquarters to the various plants and divisions or as conference-type circuits to be used by all of these parties at the same time.

In the matter of guarantee, some manufacturers offer 99.95% continuity for a 4-hop system, or close to 99.99% per hop, including all equipment failures as well as fading. These are found to be realistic and practical figures which can be attained in the field without going beyond conventional designs. It is felt, however, that it is better to avoid insistance upon an over-all performance guarantee since this leads to overdesign. Instead, it seems entirely proper to hold out for simply the conventional guarantee on materials and workmanship. Part of this comes from a realistic consideration of what would be done if performance should fall short of guarantee. It is unrealistic to conceive of removing the entire system and replacing it by one of another make, or to think of completely abandoning microwave. Logically it would be expected that the design would be rechecked and components replaced to attain the desired performance. It does not seem that this could involve a cost as great as that of overdesigning the system in the first place.

In any business the efficiency of well-planned communications combined with good operating practices is a valuable asset but for electric power system operation it becomes absolutely essential as a basic "nerve system" for the complete and continuous transmission of information from remote points to central processing centers, and from there to the generating plants. The developement of communicating facilities has kept step with the expansion and increasing complexity of power systems from the original elementary telephone systems through transmission telephone circuits, power line carrier, leased wires, long-lines telephones mobile radio, and now microwave.

Although power line carrier has been a reliable "work horse" and it is appreciated that excellent service is provided by the communications common carriers, present requirements for channels are such that these means alone are insufficient. It is very fortunate indeed that microwave has attained a level of development well beyond the experimental stage, and that the rules and regulations of the FCC permit its use in utility work. This does not mean that it or any other system would be chosen to the exclusion of others but rather that each will be exploited through what is considered the right of freedom of choice to the full extent of its technical and economic capabilities. In general, the cost of communications is in the order of 1.0% of that of the total plant and its importance is so great that there is adequate reason to lean in the direction of more and better communication facilities.

Unquestionably, in times of stress and emergency, with either potential or actual system collapse, the entire cost of the communication system is justified if it performs well and gives continuous, complete, and accurate transmission of intelligence. It is here that the ingenuity of communication engineers comes into play in designing equipment to perform under the most abnormal conditions, including complete power outage. In exploiting microwave, experience indicates that since it is capable of accomplishments not previously possible, communication engineers should be urged to show the courage and vision to plan generously for the future.

References

1. Power Supplies for Microwave Relay Systems, **H. M. Ward.** *AIEE Transactions*, vol. 68, 1949, pp. 631–36.

2. A Microwave Communication Relay System, **W. P. Boothroyd, H. J. Churchill.** *Ibid.*, pp. 637–41.

3. Microwave Applications to Bonneville Power Administration System, **R. F. Stevens, T. W. Stringfield.** *Ibid.*,vol. 69, pt. II, 1950, pp. 1088–99.

4. Field Testing a Microwave Channel for Voice Communication, Relaying, Telemetering and Supervisory Control, **D. R. Pattison, W. E. Reagan, S. C. Leyland, F. B. Gunter.** *Ibid.*, pp. 945–52.

5. Microwave Systems for 960 and 2,000 Megacycles, **R. V. Rector, W. E. Sutter.** *Ibid.*, pp. 1100–09.

6. 940–960 Megacycle Communications Equipment for Industrial Applications, **Frank B. Gunter.** *Ibid.*, vol. 70, pt. I, 1951, pp. 807–12.

7. Application of Microwave Channels, **R. C. Cheek.** *Ibid.*, pp. 813–17.

8. Protective Relaying Over Microwave Channels, **H. W. Lensner.** *Ibid.*, pt. III (*Power Apparatus and Systems*), vol. 71, Jan. 1952, pp. 240–45.

9. Microwaves in the Electric-Power Field **W. A. Derr, D. F. Burnside, H. W. Lensner.** *Ibid.*, Oct., pp. 910–17.

10. Microwave System Blankets Extensive Power Network, **H. W. Berning.** *Electric Light and Power*, Chicago, Ill., Dec. 1953.

11. Path Testing for Microwave-Radio Routes, **R. D. Campbell.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 72, July 1953, pp. 326–34.

12. Time-Division Microwave Communications for an Electric Power System, **H. D. Ashlock, I. T. Corbell.** *Ibid.*, pt. III-B (*Power Apparatus and Systems*), vol. 73, Aug. 1954, pp. 1017–27.

13. The Quadriphase Microwave System, **Sam McConoughey.** *Radio and Television News*, Chicago, Ill., Apr. 1955.

14. Recommendations for Improving Reliability of Standby-Engine Generators for Microwave Communication Systems, **N. B. Tharp.** *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 74, Apr. 1955, pp. 261–67.

15. Microwave System Parameters for Reliable Communications, **A. Alvira, I. T. Corbell.** *Ibid.*, Dec., pp. 1100–07.

16. Quadriphase—A New Approach to Time Division Multiplexing, **W. E. Evans, R. F. Lowe.** *Ibid.*, Oct., pp. 900–03.

17. A Microwave Communications System for the Southern Colorado Power Company, **G. D. Perry, Jr., Eugene Langone.** *Ibid.*, vol. 75, Oct. 1956, pp. 946–50.

18. An Approach to Microwave Relaying and System Design, **J. R. Linders, V. A. Nosko.** *Ibid.*, vol. 76, Apr. 1957, pp. 19–27.

19. A Microwave System for Protective Relaying—Boston Edison Company, **John J. Egan, Kenneth O. Sten, John E. Hurley.** *Ibid.*, Aug., pp. 585–92.

20. Field Tests of Microwave Relaying Equipment, **E. W. Downer.** *Ibid.*, Apr., pp. 74–81.

21. Evaluation of Operating Principles With Microwave Relaying, **John Berdy, H. J. Fielder, F. C. Krings, A. J. McConnel.** *Ibid.*, Oct., pp., 674–86.

◆

Discussion

G. E. Farmer (Tennessee Valley Authority, Chattanooga, Tenn.): The author is to be commended for his comprehensive and timely paper. The electric utilities are badly in need of this type of information I like the approach in respect to microwave: that it should not be expected to "assume the full burden of communications but rather should take its place to share the burden with other methods, each being applied to take maximum advantage of its individual characteristics." I should like to suggest reference 1 as an addition to the author's list.

May I comment on a few points:

Pulse-Time (PT) Versus Frequency Division. We have found PT somewhat more economical than frequency division since the latter loses the channel when it drops it at an intermediate point, whereas with PT, a new circuit may be made of the continuing section.

Signal/Noise Ratio. For us 15 db does not constitute a satisfactory circuit for one that may become a tandem circuit. *E1A* specifications recommend 35 db 97% of the time, and we use this higher figure as practically all circuits are potentially interconnectible.

Reliability. The percentage of reliability of 99.88% is remarkably good. We have not been able to approach such a figure and, as a result, do not yet consider microwave links including a repeater to be satisfactory for load control or relaying circuits. However, we are investigating the possibility of using diversity for improving this reliability figure.

Choice of Paths. We have been somewhat more conservative, using true earth's curvature on all calculations and have used full Fresnel zone clearance for paths of over 30 miles. Even with these values we still have had appreciable fading.

Stand-by Radio-Frequency. We use hot stand-by in all cases; some have 50% stand-by and others 100%, with very little difference in the relative reliability.

Emergency Power. Our propane gas engine generators have also given us excellent service. The continuous power equipment which we have used has not proved reliable and has probably caused more interruptions than it has prevented.

Channels for VHF. As our microwave repeater stations are not advantageously located, we have found it more economical to use shorter separate radio channels for this service, with voice interconnect equipment at the radio terminal.

Fading. The figure of 30 db margin is an indefinite term unless it is specific by "below the median signal or normal propagation." The industry would well be served by a more definite term.

Reference

1. Microwave—the Backbone of TVA's Expanded Communications, **G. Everett Farmer.** *Electrical World*, New York, N. Y., vol. 140, Oct. 19, 1953, pp. 124–27.

Vol.
Pt. I
1959

Functional Life Tests of Apparatus as Compared with Insulation Material Tests

M. L. MANNING
FELLOW AIEE

EXPERIMENTAL WORK by users and manufacturers of electric equipment alike, has disclosed the need for an entirely new concept of insulation classification and evaluation. It is found that insulations should be classified as systems as well as materials. The manner in which the materials are applied and processed is of as equal an importance as the basic materials used. The electrical industry, working through the AIEE, developed the concept of functional evaluation of insulation systems. This concept appears in AIEE No. *1*, published in June 1957.[1]

What does the philosophy mean to users of electric equipment?

1. It recognizes the determination of insulation life by tests made in comparison with an existing insulation system in a motor or in a transformer.
2. It necessitates the re-examination of all insulation systems used in electric machinery by a manufacturer.
3. It provides assurance that the manufacturer has taken care to provide insulation systems proved by prescribed test procedures.
4. It gives more information concerning end of life of electric equipment using certain insulation systems.

Performance of insulating materials is of utmost concern to anyone using or manufacturing electric equipment. Most troubles with electric devices can be traced to limitations of insulating materials and their inability to withstand temperature, atmospheric or environmental conditions, voltage disturbances, short circuits, and overloads. Conventional books on the subject fail to keep pace with the swift progress of research, development, and applications. Courses of study in colleges and in universities have not recognized the importance of the subject sufficiently.

This paper outlines, from the viewpoint of an educator, the need for constructive thinking today, the new concept of functional testing and what it means, the organizational procedure followed by reliable companies in making insulation studies, important properties of insulation in making an analysis, and steps involved in classifying insulating materials or systems by tests. What are the objectives of functional testing; what is the AIEE doing; what is the effect of these changes on insulation standards; how are costs of apparatus affected, and what are the possibilities for stimulating the growth of better insulating materials and insulating systems?

Definitions

The following are definitions of materials, insulation systems, and functional testing.

Insulating materials include processed compositions of raw materials and commercially available combinations before they are fabricated into structures identified with electric equipment.

Insulating systems include completely fabricated and processed combinations of insulating materials with the associated parts of specific electric equipment.

Functional or prototype testing gives the means for determining the natural characteristics and behavior of insulation systems under simulated and accelerated conditions. Insulation systems of a motor or transformer can be miniaturized in model form, then accelerated test procedures can be applied to determine insulation characteristics. Logically, this teaches that apparatus life expectancy and performance are determined, not only by what insulating materials are used, but also by environment, electrical, and mechanical stresses in an application and how they affect the life and properties of insulation.

In considering insulation as related to equipment, one factor stands out. The manufacturer is interested in the development and evaluation of an insulation system, not an insulating material by itself. It is recognized that the development of an insulation system cannot proceed in orderly fashion without the prior development of insulating materials. But the evaluation of insulating materials as materials must be considered as screening tests as far as the insulation system is concerned. Repeatedly, it has been demonstrated that evaluations must incorporate all of the processes which go into insulation systems on models which are as nearly as possible like the electric machines they simulate. For large machines, this becomes more and more difficult since the cost of models increases rapidly with size. A development program cannot overlook an insufficiently developed and evaluated system.

Organizational Procedure

The following procedures are used by reliable companies in developing insulation and corresponding systems:

1. *The research department* makes the fundamental studies in the chemistry and physics of dielectrics which result in new components.
2. *The development or long-range planning group* evaluates the materials, develops processing techniques and makes model studies; disseminates information to other departments.
3. *The application or design department* has the basic design responsibilities for applying the insulation system to equipment; must help to establish shop procedures.
4. *The service performance and maintenance department* provides the information on the insulation system in electric equipment under normal or abnormal conditions. It often recommends adjustments or changes in processing procedures.

An insulation manufacturer should basically follow the organizational procedural step numbers 1 and 2. Commercially, his success depends on steps 3 and 4, application and performance.

Properties of Insulation of Importance in Making Analysis

Mechanical: 1. Tensile strength
2. Tear resistance
3. Resistance to abrasion
4. Flexibility
5. Bond strength
6. Resiliency
7. Stability

Thermal: 1. Aging properties
2. Heat transferability
3. Temperature stability
4. Conductivity
5. Expansion and contraction

Paper 59-1293, recommended by the AIEE Electrical Insulation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE National Conference on the Application of Electrical Insulation, Washington, D. C., December 8–10, 1959. Manuscript submitted August 10, 1959; made available for printing September 29, 1959.

M. L. MANNING is with South Dakota State College of Agriculture and Mechanic Arts, Brookings, S. Dak.

Chemical: 1. Resistance to contaminants and corrosive atmospheres
2. Resistance to gases, oils, moisture, and solvents

Electrical: 1. Electric strength
2. Voltage endurance
3. Dielectric loss
4. Dielectric constant
5. Surface resistance
6. Arc resistance
7. Corona resistance
8. Radiation resistance
9. Power factor

Classifying Insulating Materials

What steps are involved in classifying insulating materials or systems by tests?

1. A thermal test procedure may be proposed by an individual or a committee.
2. Temperature class limits are set by tests on materials or systems.
3. Limiting test values are selected for each temperature.
4. A committee may accept a proposed procedure, organize a test program, and publish results.
5. A test procedure may be published as a standard.

Objectives

Following are the objectives of functional testing:

1. To offer suggestions for designing test samples or models.
2. To determine heat-cycling limits, environmental conditions, voltage or end-point tests for developing test procedures.
3. To determine life expectancy of insulation or of insulation systems (life versus temperature relations).
4. To determine the mechanical, thermal, chemical, and electrical properties outlined previously.
5. To determine the effects of high temperature under different environments on the insulation or the system.

What Is the AIEE Doing About the Problem?

Insulating Materials

The AIEE Standards Co-ordinating Committee No. *4* has prepared Test Procedure No. *1D*[2] in which principles are established for the development of test procedures for the thermal evaluation of materials. A test procedure is prepared for each important kind of insulation material such as wire enamel,[3] rigid insulating materials, sheet insulation, and other components.

Materials are assigned to a temperature class to facilitate comparison. It must be remembered that life expectancy under the test conditions chosen may not have uniform relation to the material in actual service. The insulation system in which the material is to be used may have widely different temperature limits.

Insulation Systems

AIEE Committee No. *4* has published a "Guide for Preparing Test Procedures" which provides classification of insulation systems for electric equipment on a temperature basis by test, rather than by chemical composition of the component materials of the system.[4]

Statistical Analysis of Test Data

AIEE No. *1* states that accelerated life tests are recommended to determine the probable life expectancy of insulating materials and of complete insulation systems under different conditions of service.

After the life data from tests are obtained, the problem is to evaluate such data so that the life expectancy of insulation can be estimated. "A Guide for the Statistical Analysis of Test Data" is now available.[5]

Test Procedures for Electric Equipment

For motors, generators, and transformers, several test procedures are available.[6–8] These procedures serve the purpose of accelerating insulation degradation at various temperatures.

Effect of AIEE No. 1 Revisions on Insulating Standards[9–11]

Insulation System Tests Using Models

1. To begin the testing process, simulate the actual insulation system and, if practical, use actual equipment.
2. Provision is made for a composite picture by procedures for short as well as long-time voltage-endurance tests.
3. Thermal aging properties are better determined by a combination of stresses on the complete insulation system at room and at elevated temperatures.
4. Processing procedures are a duplicate of those on equipment and can be reduced to practice.
5. Mechanical properties, such as bond strength, shear strength, compressive strength, and creep properties are obtainable at operating temperatures.
6. Progress is made by steps so that correlation of test results with those found on basic materials is possible.
7. Test procedures recognize degrading effect of associated materials on the system and consideration is given to the decomposition products.
8. Effects of temperature are understood more clearly and lead to the design of equipment for higher temperatures in order to reduce space and weight.
9. Life tests can be accelerated at high temperatures and at high mechanical stresses; high voltage can be used to determine creepage strength; and the effects of corona can be studied.
10. Correlation with machine standards is possible.

Insulating Materials

1. It is impossible to design test samples to simulate actual usage conditions.
2. Insulating materials are usually designed without known end-point use.
3. It is impossible to create electrical, thermal, and mechanical stresses found in equipment.
4. Materials must be evaluated by "screen-testing" methods.
5. Correlation of test results with those found on insulation systems must be made.
6. Tests may predict temperature limits for materials to determine temperature classes.
7. Co-operative work with the American Society of Testing Materials is possible to establish test standards.
8. Development of test procedures must recognize: (*a*). preliminary plans made by experts in materials; and (*b*). the establishment of tests such that they are reproducible and that a reasonable correlation between test life and service life is possible.

Vol.
Pt. 1
195

How Are Costs of Apparatus Affected?

The AIEE guides, in their philosophy, offer one of the best quality control devices given to a user of electric equipment. They demand that both the manufacturer of insulating materials and of electric machinery must use high-quality materials and should display evidence of functional testing of the complete insulation system. Manufacturers of higher temperature materials will be benefited in the long run. Temporarily, for lower temperature applications, the demand may now be greater for organic insulation materials. As industry progresses and as pressure is generated for space and weight reduction and the minimizing of fire hazards becomes more pronounced, the high-temperature-withstanding materials will be in greater demand.

It seems logical that a purchaser of equipment should take these factors into consideration. Companies who take active part in AIEE committee activities by conducting quite expensive laboratory investigations and tests should be recognized and some method for prorating these costs should be devised. Logically, it would seem that these extra considerations should be factored additionally in arriving at the cost of a product.

Possibilities for Better Insulating Materials and Systems

Without doubt, the complexities of the many forms of insulation and corresponding usage in systems can be sorted out and verified by the new concept of functional testing.[12–23] Research and development demonstrate that chosen temperature limits have many, many facets, such as:

1. Effects of long exposure to high temperature. Do the materials change with respect to mechanical and electrical properties? Do the materials soften, lose weight, or become brittle?

2. The magnitude and time duration of loading imposed on the insulation system.

3. The effect on the physical support of insulating materials at high temperatures. The bonding resins may become brittle, yet the insulation system may continue to function if not disturbed. Forces of thermal expansion and contraction may restrict temperature limitations depending upon the size of electric equipment.

4. Life of insulation is dependent on environment, such as moisture, dirt, chemicals, or solvents, released from the interior of an insulation system. Various industrial atmospheres may have a predominant influence on ventilated apparatus compared with sealed systems which exclude oxidation or contamination.

5. The rate of physical deterioration increases rapidly as temperature increases. The modern concept of determining insulation life at elevated temperature is based upon the idea that the logarithm of insulation life is a function of the reciprocal of the absolute temperature. At three or more temperatures the logarithm of the hours of life found by thermal evaluation tests are plotted against the reciprocal of the absolute temperatures. Usually a straight line results. The life at other temperatures can be estimated. Individual time-temperature life curves for different materials or insulation systems can be determined by thermal evaluation tests.[24]

   Basically, the idea that electrical insulation deterioration is treated as a chemical rate phenomenon stems from the Arrhenius chemical reaction rate equation.[24]

6. Ambient temperature affects the temperature attained by equipment in operation.

7. Insulation life depends on the care and maintenance of equipment on which it is used. Mechanical shocks, such as movement of equipment to different locations, short circuits, and overvoltage, are factors to consider.

In general, a method is needed to grind out, like a mill, functional merit ratings of insulation systems. All of the factors considered may be likened to seven prongs impinging on insulation simultaneously.

A tremendous amount of work lies before the industry. The job cannot be done overnight. Years of demanding co-operative effort between users and manufacturers are required to inaugurate AIEE test procedures, to realize that ratings of machines must be on a temperature basis according to the insulation system within them, and to establish temperature ratings under normal conditions with accelerated test conditions. Excessive time is required at normal temperature conditions to establish life relationships. Accelerated tests may be used to expedite better answers than those now available.

Users of electric equipment can be of great assistance by becoming members of various AIEE Committees and taking active part in establishing test procedures, and by keeping accurate records of performance of equipment subject to overloads or to abnormal environmental conditions.

A 3-dimensional relationship of time (duration of load), magnitude of load, and life of insulation system for apparatus is the answer desired. Such a relationship, resulting in a 3-dimensional graph in terms of per-cent load, will sort out answers to amortization of life of equipment when applications demand 2 hours, 3 years, 6 years, 10 years, or 30 years of life.

Co-operative papers written by electric power representatives and manufacturers are now appearing in publications.[25–27] By actual load charts, by record of environmental conditions, by a careful check of mortality rates on equipment, and by co-ordinated effort, the aims expressed in this paper can be realized.

In this day of space, rockets, guided missiles, and satellites, the "triple alliance" effort of physicists, engineers, and chemists in the field of electrical insulation, will develop new insulation and systems far better than those now known. The past three decades of progress in electrical insulation are quite notable[28] but there is no room for complacency. Unanswered questions concerning insulation and systems, proposed 8 years ago at the National Conference on Electrical Insulation, remain with us.

1. What magnitude of test voltages should be used for end-point tests? What about corona damage at higher voltages?

2. What about the detection of corona in the volume as well as over the surface of an insulation system?

3. What about chemical changes which take place, thereby giving off by-products which affect rate of reaction on other materials in the system?

4. What electrical and mechanical performance is required for a given application?

5. What time versus load duty cycle is expected? Are sufficient data available?

6. What life expectancy and reliability can be realized?

References

1. General Principles Upon Which Temperature Limits Are Based in the Rating of Electric Equipment. *AIEE No. 1*, revised June 1957.

2. Guide for the Preparation of Test Procedures for the Thermal Evaluation of Electrical Insulating Materials. *AIEE No. 1D*, June 1957.

3. Proposed Test Procedure for Evaluation of the Thermal Stability of Enameled Wire in Air. *AIEE Standard no. 57*, Oct. 1955.

4. Guide for the Preparation of Test Procedures for the Thermal Evaluation of Insulation Systems for Electric Equipment. *AIEE No. 1E*, June 1957.

5. Guide for Statistical Analysis of Test Data. *AIEE No. 1F*, June 1958.

6. Test Procedure for Evaluation of Systems of Insulating Materials for Random-Wound Electric Machinery. *AIEE Standard no. 51*, Nov. 1956.

7. Proposed Test Procedure for Evaluation of Systems of Insulating Materials for Electric Machinery Employing Form-Wound Pre-Insulated Coils. *AIEE Publication no. 511*, Oct. 1956.

8. Proposed Test Procedure for Thermal Evaluation of Ventilated Dry-Type Power and Distribution Transformers. *AIEE Publication no. 65*, Nov. 1956.

9. Thermal Endurance Tests for Oil-Filled Distribution Transformers. (AIEE test procedure now in process.)

10. A Bibliography on Testing of Insulating Materials and Systems for Thermal Degradation. *AIEE Special Publication S-87*, Apr. 1956.

11. The Development of Standards for Insulation Systems and Insulating Materials, M. L. Schmidt. *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 77, Oct. 1958, pp. 747–49.

12. New Concepts in Evaluating Electric Equipment Insulation Systems—The Problems Encountered, M. L. Manning. *Ibid.*, pt. I (*Communication and Electronics*), vol. 76, July 1957, pp. 348–55.

13. Encyclopedia of Electrical Insulation, Gerald de Senarclens. *Electrical Manufacturing*, New York, N. Y., vol. 58, no. 6, Dec. 1956, pp. 92–99.

14. Functional Temperature Endurance Tests on a Silicone Glass Fiber Insulation System for Dry-Type Transformers, M. L. Manning. *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 74, Apr. 1955, pp. 91–100.

15. A Study of Models for Use in Evaluating Dry-Type Transformer Insulating Systems, T. R. Walters, A. L. Scheideler. *Ibid.*, vol. 75, Aug. 1956, pp. 520–27.

16. Functional Life Tests of Insulation Systems for Dry-Type Transformers, Paul Narbut. *Ibid.*, vol. 74, Dec. 1955, pp. 1057–62.

17. Functional Life Evaluation of Group-2 Dry-Type Power Transformers, H. G. Zambell. *Ibid.*, Oct., pp. 956–70.

18. The Functional Evaluation of Insulation for Small Dry-Type Transformers Used in Electronic Equipment, R. L. Hamilton, H. B. Harms. *Ibid.*, vol. 73, June 1954, pp. 605–10.

19. Motor Insulation Life as Measured by Accelerated Tests and Dielectric Fatigue,

C. J. Herman. *Ibid.*, vol. 72, Oct. 1953, pp. 986–91.

20. Insulation Systems for Random-Wound Motors Evaluated by Motorette Tests, T. J. Gair. *Ibid.*, vol. 73, 1954 (Feb. 1955 section), pp. 1702–07.

21. Functional Life Expectancy Tests for Liquid-Filled Distribution Transformers, A. M. Lockie. *Ibid.*, vol. 74, Oct. 1955, pp. 977–85.

22. Accelerated Aging Characteristics of Formex- and Paper-Insulated Wires in Transformer Oil, **M. F. Beavers**, H. H. Brustle, J. H. Carpenter, W. J. Degnan. *Ibid.*, vol. 78, Dec. 1959, pp. 1202–07.

23. Effect of Coil Winding and Processing on the Electrical Strength of Oil-Immersed Insulating Materials, **W. T. Sackett, Jr.** *Ibid.*, vol. 77, Apr. 1958, pp. 118–23.

24. Electrical Insulation Deterioration Treated as a Chemical Rate Phenomenon, **Thomas W. Dakin.** *Ibid.*, vol. 67, pt. I, 1948, pp. 113–22.

25. Economic Loading of Distribution Transformers, **C. F. Mitchell.** *Proceedings*, American Power Conference, Chicago, Ill., Mar. 21, 1956.

26. An Economic Analysis of Distribution Transformer Application, **C. F. Mitchell, J. O. Sweeny, J. L. Cantwell.** *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 78, Dec. 1959, pp. 1196–1202.

27. Pole-Mounted Distribution Transformers of the Future, **A. J. Pansini, M. L. Manning.** Edison Electric Institute, Chicago, Ill., May 1956.

28. Three Decades of Progress in Electrical Insulation, **L. J. Berberich.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, July 1959, pp. 237–48.

Discussion

H. P. Bruncke (Northern States Power Company, Minneapolis, Minn.): Dean Manning has compiled in his relatively short thought-provoking paper, a world of information concerning tests of electric apparatus. As he has pointed out, apparatus life is dependent not upon the life of the separate materials in the apparatus but upon the life of the insulation system used for this apparatus. The organization of the paper gives much food for thought on the many aspects of the insulation system. As we all know and have experienced, tests in the factory or the laboratory can never duplicate the actual field service conditions. There is always some factor which has been either overlooked or which was not apparent in the design of the tests for the apparatus.

In the past there has been the tendency for manufacturers of insulation systems to depend too much upon tests of the individual materials suppliers. Now the AIEE has amplified AIEE No. *1* to take into consideration the system rather than the individual materials going into the system. Simulated or accelerated tests on materials can never take into consideration the operating conditions of the insulating system.

Both time and ambient temperatures have a very real effect upon the life of insulation. In all testing conditions, emphasis has been placed upon temperature; generally, the higher temperatures. The assumption has been, of course, that high temperatures reduce the life of insulating materials. There is a feeling among the users of electric equipment that not only high temperatures but also low temperatures are very real factors to include in any standard analysis. As an example, we had several cases of trouble with a compound-filled 15-kv class potential transformer. These transformers failed after a period of very cold weather. The failures seemed to be limited to our section of the country. An analysis indicated that during the very cold weather, because of the small heat loss, the compound cracked setting up a corona condition which developed as the ambient temperature increased. In some cases the instrument transformers failed with explosive force. This example is cited to indicate that not only tests for life regarding temperature should be included, but also tests under operating conditions should be included. The AIEE's No. *1* also includes weather conditions in the United States giving maximums, minimums, and averages which give the laboratory much information for operating conditions.

Dean Manning has recommended that the users of electric equipment assist this testing procedure by becoming members of the various AIEE committees and by furnishing records showing normal and abnormal operating conditions. This paragraph of Dean Manning's paper brought to my mind remarks made by the chairman of the AIEE Transformer Committee about a year ago in which he lamented the fact that there were too few members on his Committee from the utilities or the industrial group; that is, the users. Nearly all of the committee members represented the manufacturers. The manufacturers need the information of the user in order to correlate test data.

Dean Manning has offered a challenge to a number of unanswered questions for insulation and insulation systems. He has enumerated those which were proposed at the National Conference on Electrical Insulation which apparently have been left unanswered. One question which the utility users are very much interested in is the detection of corona in insulating materials as well as over the surface of an insulation system. There has been more and more evidence in the analyses of failures that corona is a very real problem. Our usual test methods in the field are not sensitive enough for detection at the start of a corona condition. Apparently the only symptoms of corona problems in insulation are radio, television, or carrier-current interference. When this occurs, then there are methods for isolating the defective piece of equipment. There is a need to detect this corona condition prior to this time.

Dean Manning is to be congratulated on his paper and for the excellent arrangement which brings to our minds the various problems which are factors in the production of satisfactory electric apparatus.

P. L. Alger (Schenectady, N. Y.): I agree thoroughly with the statements made in this paper. Evidently the whole industry will have to go on working in the evaluation field for years to come, and it is a real problem how to select and guide the tasks that each group should perform.

It seems to me that a chief factor in future evaluation work will be to study and control the effects of corona. For, in order to use the new and more expensive insulating materials to advantage it is essential that they be made thinner. This increases the dielectric stress and, therefore, the risk of corona damage. Finding materials that will not be damaged by corona, such as silicone rubber, is one way of progress. Another way is to find means to prevent the occurrence of corona by shielding or environmental control. In either case, methods are needed for quickly evaluating corona effects.

In my opinion, an important step in this direction was recorded in the paper by Rhudy and Mazanek, in which they showed that the use of high frequency, 3,000 cycles, for example, gave results fully comparable to 60-cycle tests on corona damage to organic insulated armature coils, in a small fraction of the time. I regret, therefore, that the author did not mention this article in his extensive bibliography.

Reference

1. Accelerated Voltage Endurance Tests, **R. G. Rhudy, H. E. Mazanek.** *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 79, 1960 (*Paper no. 58-1391*).

P. H. Ware and **D. W. Kitchin**[*] (Simplex Wire & Cable Company, Cambridge, Mass.): Mr. Manning has made a very thorough study of the philosophy underlying the use of testing as an aid to and a verification of engineering design. As he points out, there is a trend away from testing of insulating material per se, and toward the testing of insulating systems.

While we appreciate the fact that insulation when incorporated into systems does not always behave in the same way as the insulation when tested in a form convenient for quick evaluation, we feel that there may be appreciable loss in departing from long-established laboratory practice. The insulation manufacturer, as opposed to the equipment manufacturer, will be severely handicapped if he is to rely on system testing alone. Much of the repetitive testing must be done even to find the materials which show promise. If Thomas Edison had not used methods of quick evaluation for materials, much of his work would have been greatly delayed.

In general, testing of insulation systems in the form of models is more expensive and takes longer to perform. The necessary screening must be done as quickly as possible to find the materials worth putting into model form. The need for sufficient lead time between design and production often precludes a long-drawn-out program of simulated service tests.

One point of which we were reminded by the paper is that accelerated tests do not always produce results which bear a true relationship to service operation. For example, if acceleration is accomplished by using higher than normal voltages, life may be shortened by the effects of corona which does not exist at operating voltage.

Similarly, acceleration using increased temperature may cause a change in state which then will invalidate the entire test.

Basically, we believe that a balanced mixture of the two philosophies is required. To determine the suitability of fundamental properties, screening tests can be devised so that each necessary property of a wide range of materials may be evaluated. Then, once the obviously unfit materials have been weeded out, a carefully designed experiment considering all variables may be performed. In this latter test, the simulation of the actual system can be considered, and the model system may be made as nearly like the equipment it represents as is economically desirable.

In fact, Mr. Manning points out in his discussion of organizational procedure followed by reliable companies that just this balance of material tests and system tests takes place. The paper brings out many facets of the insulation testing problem.

M. L. Manning: The author appreciates the comments of the discussers.

Mr. Bruncke pointed out that time and ambient temperatures have a real effect upon the life of insulation. He illustrates the fact that low ambient temperature is a factor to be considered in compound-filled potential transformers. This amplifies the idea that more users of electric equipment should be members of various AIEE Committees dealing with electrical insulation. The importance of corona detection, as explained by Mr. Bruncke, is interesting.

Mr. Alger brings out an important consideration in the study and control of the effects of corona. Thinner materials, to reduce space factors, will be used in the future. This results in higher electric stresses and resultant corona. Means must be found to prevent corona formation and to find methods to detect it.

Mr. Ware and Mr. Kitchin explain that insulation systems should not be overemphasized in importance, compared with insulation materials alone. This is true, but screen tests on materials should be made first, in line with AIEE procedures. Uniform standard tests should be considered, not a battery of tests made at the whims of a manufacturer. The purpose of the AIEE program is to establish a balance between insulation materials per se and insulation systems in which they are used. As the years go by, it is hoped this philosophy gains much more momentum. As the two discussers point out, reliable companies are in accord with the organizational procedure described in the paper.

It is hoped the aims of this paper will be carried out to much greater extent in the decade before us. Co-operation among all concerned is highly important.